

# Posterior concentration rates for counting processes with Aalen multiplicative intensities

Sophie Donnet <sup>\*</sup> and Vincent Rivoirard <sup>†</sup> and Judith Rousseau <sup>‡</sup> and Catia Scricciolo <sup>§</sup>

**Abstract.** We provide general conditions to derive posterior concentration rates for Aalen counting processes. The conditions are designed to resemble those proposed in the literature for the problem of density estimation, for instance in Ghosal et al. (2000), so that existing results on density estimation can be adapted to the present setting. We apply the general theorem to some prior models including Dirichlet process mixtures of uniform densities to estimate monotone non-increasing intensities and log-splines.

**Keywords:** Aalen model, counting processes, Dirichlet process mixtures, posterior concentration rates

## 1 Introduction

Estimation of the intensity function of a point process is an important statistical problem with a long history. Most methods were initially employed for estimating intensities assumed to be of parametric or nonparametric form in Poisson point processes. However, in many fields such as genetics, seismology and neuroscience, the probability of observing a new occurrence of the studied temporal process may depend on covariates and, in this case, the intensity of the process is random so that such a feature is not captured by a classical Poisson model. Aalen models constitute a natural extension of Poisson models that allow taking into account this aspect. Aalen (1978) revolutionized point processes analysis developing a unified theory for frequentist nonparametric inference of multiplicative intensity models which, besides the Poisson model and other classical models such as right-censoring and Markov processes with finite state space, described in Section 1.1, encompass birth and death processes as well as branching processes. We refer the reader to Andersen et al. (1993) for a presentation of Aalen processes including various other illustrative examples. Classical probabilistic and statistical results about Aalen processes can be found in Karr (1991), Andersen et al. (1993), Daley and Vere-Jones (2003, 2008). Recent nonparametric frequentist methodologies based on penalized least-squares contrasts have been proposed by Brunel and Comte (2005, 2008), Comte et al. (2011) and Reynaud-Bouret (2006). In the high-dimensional setting, more specific results have been established by Gaïffas and Guillaoux (2012) and Hansen et al. (2012) who consider Lasso-type procedures.

Bayesian nonparametric inference for inhomogeneous Poisson point processes has

---

<sup>\*</sup>MIA. INRA. UMR0518, AgroParisTech, donnet@cimat.mx

<sup>†</sup>CEREMADE, Université Paris Dauphine, rivoirard@ceremade.dauphine.fr

<sup>‡</sup>ENSAE-CREST, rousseau@ceremade.dauphine.fr

<sup>§</sup>Department of Decision Sciences, Bocconi University, catia.scriciolo@unibocconi.it

been considered by Lo (1982) who develops a prior-to-posterior analysis for weighted gamma process priors to model intensity functions. In the same spirit, Kuo and Ghosh (1997) employ several classes of nonparametric priors, including the gamma, the beta and the extended gamma processes. Extension to multiplicative counting processes has been treated in Lo and Weng (1989), who model intensities as kernel mixtures with mixing measure distributed according to a weighted gamma measure on the real line. Along the same lines, Ishwaran and James (2004) develop computational procedures for Bayesian non- and semi-parametric multiplicative intensity models using kernel mixtures of weighted gamma measures. Other papers have mainly focussed on exploring prior distributions on intensity functions with the aim of showing that Bayesian nonparametric inference for inhomogeneous Poisson processes can give satisfactory results in applications, see, e.g., Kottas and Sansó (2007).

Surprisingly, leaving aside the recent work of Belitser et al. (2013), which deals with optimal convergence rates for estimating intensities in inhomogeneous Poisson processes, there are no results in the literature concerning aspects of the frequentist asymptotic behaviour of posterior distributions, like consistency and rates of convergence, for intensity estimation of general Aalen models. In this paper, we extend their results to general Aalen multiplicative intensity models. Quoting Lo and Weng (1989), “the idea of our approach is that estimating a density and estimating a hazard rate are analogous affairs, and a successful attempt of one generally leads to a feasible approach for the other”. Thus, in deriving general sufficient conditions for assessing posterior contraction rates in Theorem 2.1 of Section 2, we attempt at giving conditions which resemble those proposed by Ghosal et al. (2000) for density estimation with independent and identically distributed (i.i.d.) observations. This allows us to then derive in Section 3 posterior contraction rates for different families of prior distributions, such as Dirichlet mixtures of uniform densities to estimate monotone non-increasing intensities and log-splines, by an adaptation of existing results on density estimation. Detailed proofs of the main results are reported in Section 4. Auxiliary results concerning the control of the Kullback-Leibler divergence for intensities in Aalen models and existence of tests, which, to the best of our knowledge, are derived here for the first time and can also be of independent interest, are presented in Section 5 and in Section 6.

## 1.1 Notation and set-up

We observe a counting process  $N$  and denote by  $(\mathcal{G}_t)_t$  its adapted filtration. Let  $\Lambda$  be the compensator of  $N$ . We assume it satisfies the condition  $\Lambda_t < \infty$  almost surely for every  $t$ . Recall that  $(N_t - \Lambda_t)_t$  is a zero-mean  $(\mathcal{G}_t)_t$ -martingale. We assume that  $N$  obeys the *Aalen multiplicative intensity model*

$$d\Lambda_t = Y_t \lambda(t) dt,$$

where  $\lambda$  is a non-negative deterministic function called *intensity function* in the sequel and  $(Y_t)_t$  is a non-negative predictable process. Informally,

$$\mathbb{P}[N[t, t + dt] \geq 1 \mid \mathcal{G}_{t-}] = Y_t \lambda(t) dt, \quad (1.1)$$

see Andersen et al. (1993), Chapter III. In this paper, we are interested in asymptotic results: both  $N$  and  $Y$  depend on an integer  $n$  and we study estimation of  $\lambda$  (not depending on  $n$ ) when  $T$  is kept fixed and  $n \rightarrow \infty$ . The following special cases motivate the interest in this model.

### Inhomogeneous Poisson processes

We observe  $n$  independent Poisson processes with common intensity  $\lambda$ . This model is equivalent to the model where we observe a Poisson process with intensity  $n \times \lambda$ , so it corresponds to the case  $Y_t \equiv n$ .

### Survival analysis with right-censoring

This model is popular in biomedical problems. We have  $n$  patients and, for each patient  $i$ , we observe  $(Z_i, \delta_i)$ , with  $Z_i = \min\{X_i, C_i\}$ , where  $X_i$  represents the lifetime of the patient,  $C_i$  is the independent censoring time and  $\delta_i = \mathbf{1}_{X_i \leq C_i}$ . In this case, we set  $N_t^i = \delta_i \times \mathbf{1}_{Z_i \leq t}$ ,  $Y_t^i = \mathbf{1}_{Z_i \geq t}$  and  $\lambda$  is the hazard rate of the  $X_i$ 's: if  $f$  is the density of  $X_1$ , then  $\lambda(t) = f(t)/\mathbb{P}(X_1 \geq t)$ . Thus,  $N$  (respectively  $Y$ ) is obtained by aggregating the  $n$  independent processes  $N^i$ 's (respectively the  $Y^i$ 's): for any  $t \in [0, T]$ ,  $N_t = \sum_{i=1}^n N_t^i$  and  $Y_t = \sum_{i=1}^n Y_t^i$ .

### Finite state Markov processes

Let  $X = (X(t))_t$  be a Markov process with finite state space  $\mathbb{S}$  and right-continuous sample paths. We assume the existence of integrable *transition intensities*  $\lambda_{hj}$  from state  $h$  to state  $j$  for  $h \neq j$ . We assume we are given  $n$  independent copies of the process  $X$ , denoted by  $X^1, \dots, X^n$ . For any  $i \in \{1, \dots, n\}$ , let  $N_t^{ihj}$  be the number of direct transitions for  $X^i$  from  $h$  to  $j$  in  $[0, t]$ , for  $h \neq j$ . Then, the intensity of the multivariate counting process  $\mathfrak{N}^i = (N^{ihj})_{h \neq j}$  is  $(\lambda_{hj} Y^{ih})_{h \neq j}$ , with  $Y_t^{ih} = \mathbf{1}_{\{X^i(t^-)=h\}}$ . As before, we can consider  $\mathfrak{N}$  (respectively  $Y^h$ ) by aggregating the processes  $\mathfrak{N}^i$  (respectively the  $Y^{ih}$ 's):  $\mathfrak{N}_t = \sum_{i=1}^n \mathfrak{N}_t^i$ ,  $Y_t^h = \sum_{i=1}^n Y_t^{ih}$  and  $t \in [0, T]$ . The intensity of each component  $(N_t^{hj})_t$  of  $(\mathfrak{N}_t)_t$  is then  $(\lambda_{hj}(t) Y_t^h)_t$ . We refer the reader to Andersen et al. (1993), p. 126, for more details. In this case,  $N$  is either one of the  $N^{hj}$ 's or the aggregation of some processes for which the  $\lambda_{hj}$ 's are equal.

We now state some conditions concerning the asymptotic behavior of  $Y_t$  under the true intensity function  $\lambda_0$ . Define  $\mu_n(t) := \mathbb{E}_{\lambda_0}^{(n)}[Y_t]$  and  $\tilde{\mu}_n(t) := n^{-1}\mu_n(t)$ . We assume the existence of a non-random set  $\Omega \subseteq [0, T]$  such that there are constants  $m_1$  and  $m_2$  satisfying

$$m_1 \leq \inf_{t \in \Omega} \tilde{\mu}_n(t) \leq \sup_{t \in \Omega} \tilde{\mu}_n(t) \leq m_2 \quad \text{for every } n \text{ large enough,} \quad (1.2)$$

and there exists  $\alpha \in (0, 1)$  such that, if  $\Gamma_n := \{\sup_{t \in \Omega} |n^{-1}Y_t - \tilde{\mu}_n(t)| \leq \alpha m_1\} \cap \{\sup_{t \in \Omega^c} Y_t = 0\}$ , where  $\Omega^c$  is the complement of  $\Omega$  in  $[0, T]$ , then

$$\lim_{n \rightarrow \infty} \mathbb{P}_{\lambda_0}^{(n)}(\Gamma_n) = 1. \quad (1.3)$$

We only consider estimation over  $\Omega$  ( $N$  is almost surely empty on  $\Omega^c$ ) and define the parameter space as  $\mathcal{F} = \{\lambda : \Omega \rightarrow \mathbb{R}_+ \mid \int_{\Omega} \lambda(t) dt < \infty\}$ . Let  $\lambda_0 \in \mathcal{F}$ .

For inhomogeneous Poisson processes, conditions (1.2) and (1.3) are trivially satisfied for  $\Omega = [0, T]$  since  $Y_t \equiv \mu_n(t) \equiv n$ . For right-censoring models, with  $Y_t^i = \mathbf{1}_{Z_i \geq t}$ ,  $i = 1, \dots, n$ , we denote by  $\Omega$  the support of the  $Z_i$ 's and by  $M_\Omega = \max \Omega \in \mathbb{R}_+$ . Then, (1.2) and (1.3) are satisfied if  $M_\Omega > T$  or  $M_\Omega \leq T$  and  $\mathbb{P}(Z_1 = M_\Omega) > 0$  (the concentration inequality is implied by an application of the DKW inequality).

We denote by  $\|\cdot\|_1$  the  $\mathbb{L}_1$ -norm over  $\mathcal{F}$ : for all  $\lambda, \lambda' \in \mathcal{F}$ ,  $\|\lambda - \lambda'\|_1 = \int_\Omega |\lambda(t) - \lambda'(t)| dt$ .

## 2 Posterior contraction rates for Aalen counting processes

In this section, we present the main result providing general sufficient conditions for assessing concentration rates of posterior distributions of intensities in general Aalen models. Before stating the theorem, we need to introduce some more notation.

For any  $\lambda \in \mathcal{F}$ , we introduce the following parametrization  $\lambda = M_\lambda \times \bar{\lambda}$ , where  $M_\lambda = \int_\Omega \lambda(t) dt$  and  $\bar{\lambda} \in \mathcal{F}_1$ , with  $\mathcal{F}_1 = \{\lambda \in \mathcal{F} : \int_\Omega \lambda(t) dt = 1\}$ . For the sake of simplicity, in this paper we restrict attention to the case where  $M_\lambda$  and  $\bar{\lambda}$  are *a priori* independent so that the prior probability measure  $\pi$  on  $\mathcal{F}$  is the product measure  $\pi_1 \otimes \pi_M$ , where  $\pi_1$  is a probability measure on  $\mathcal{F}_1$  and  $\pi_M$  is a probability measure on  $\mathbb{R}_+$ . Let  $v_n$  be a positive sequence such that  $v_n \rightarrow 0$  and  $nv_n^2 \rightarrow \infty$ . For every  $j \in \mathbb{N}$ , we define

$$\bar{S}_{n,j} = \{\bar{\lambda} \in \mathcal{F}_1 : \|\bar{\lambda} - \bar{\lambda}_0\|_1 \leq 2(j+1)v_n/M_{\lambda_0}\},$$

where  $M_{\lambda_0} = \int_\Omega \lambda_0(t) dt$  and  $\bar{\lambda}_0 = M_{\lambda_0}^{-1} \lambda_0$ . For  $H > 0$  and  $k \geq 2$ , if  $k_{[2]} = \min\{2^\ell : \ell \in \mathbb{N}, 2^\ell \geq k\}$ , we define

$$\bar{B}_{k,n}(\bar{\lambda}_0; v_n, H) = \left\{ \bar{\lambda} \in \mathcal{F}_1 : h^2(\bar{\lambda}_0, \bar{\lambda}) \leq v_n^2 / (1 + \log \|\bar{\lambda}_0 / \bar{\lambda}\|_\infty), \right. \\ \left. \max_{2 \leq j \leq k_{[2]}} E_j(\bar{\lambda}_0; \bar{\lambda}) \leq v_n^2, \|\bar{\lambda}_0 / \bar{\lambda}\|_\infty \leq n^H, \|\bar{\lambda}\|_\infty \leq H \right\},$$

where  $h^2(\bar{\lambda}_0, \bar{\lambda}) = \int_\Omega (\sqrt{\lambda_0(t)} - \sqrt{\lambda(t)})^2 dt$  is the squared Hellinger distance between  $\bar{\lambda}_0$  and  $\bar{\lambda}$ ,  $\|\cdot\|_\infty$  stands for the sup-norm and  $E_j(\bar{\lambda}_0; \bar{\lambda}) := \int_\Omega \bar{\lambda}_0(t) |\log \bar{\lambda}_0(t) - \log \bar{\lambda}(t)|^j dt$ . In what follows, for any set  $\Theta$  equipped with a semi-metric  $d$  and any real number  $\epsilon > 0$ , we denote by  $D(\epsilon, \Theta, d)$  the  $\epsilon$ -packing number of  $\Theta$ , that is, the maximal number of points in  $\Theta$  such that the  $d$ -distance between every pair is at least  $\epsilon$ . Since  $D(\epsilon, \Theta, d)$  is bounded above by the  $(\epsilon/2)$ -covering number, namely, the minimal number of balls of  $d$ -radius  $\epsilon/2$  needed to cover  $\Theta$ , with abuse of language, we will just speak of covering numbers. We denote by  $\pi(\cdot | N)$  the posterior distribution of the intensity function  $\lambda$ , given the observations of the process  $N$ .

**Theorem 2.1.** *Assume that conditions (1.2) and (1.3) are satisfied and that, for some  $k \geq 2$ , there exists a constant  $C_{1k} > 0$  such that*

$$\mathbb{E}_{\lambda_0}^{(n)} \left[ \left( \int_\Omega [Y_t - \mu_n(t)]^2 dt \right)^k \right] \leq C_{1k} n^k. \quad (2.1)$$

Assume that the prior  $\pi_M$  on the mass  $M$  is absolutely continuous with respect to Lebesgue measure and has positive and continuous density on  $\mathbb{R}_+$ , while the prior  $\pi_1$  on  $\bar{\lambda}$  satisfies the following conditions for some constant  $H > 0$ :

- (i) there exists  $\mathcal{F}_n \subseteq \mathcal{F}_1$  such that, for a positive sequence  $v_n = o(1)$  and  $v_n^2 \geq (n/\log n)^{-1}$ ,

$$\pi_1(\mathcal{F}_n^c) \leq e^{-(\kappa_0+2)nv_n^2} \pi_1(\bar{B}_{k,n}(\bar{\lambda}_0; v_n, H)),$$

with

$$\kappa_0 = m_2^2 M_{\lambda_0} \left\{ \frac{4}{m_1} \left[ 1 + \log \left( \frac{m_2}{m_1} \right) \right] \left( 1 + \frac{m_2^2}{m_1^2} \right) + \frac{m_2(2M_{\lambda_0} + 1)^2}{m_1^2 M_{\lambda_0}^2} \right\}, \quad (2.2)$$

and, for any  $\xi, \delta > 0$ ,

$$\log D(\xi, \mathcal{F}_n, \|\cdot\|_1) \leq n\delta \quad \text{for all } n \text{ large enough;}$$

- (ii) for all  $\zeta, \delta > 0$ , there exists  $J_0 > 0$  such that, for every  $j \geq J_0$ ,

$$\frac{\pi_1(\bar{S}_{n,j})}{\pi_1(\bar{B}_{k,n}(\bar{\lambda}_0; v_n, H))} \leq e^{\delta(j+1)^2 nv_n^2}$$

and

$$\log D(\zeta j v_n, \bar{S}_{n,j} \cap \mathcal{F}_n, \|\cdot\|_1) \leq \delta(j+1)^2 nv_n^2.$$

Then, there exists a constant  $J_1 > 0$  such that

$$\mathbb{E}_{\lambda_0}^{(n)}[\pi(\lambda : \|\lambda - \lambda_0\|_1 > J_1 v_n \mid N)] = O((nv_n^2)^{-k/2}).$$

The proof of Theorem 2.1 is reported in Section 4. To the best of our knowledge, the only other paper dealing with posterior concentration rates in related models is that of Belitser et al. (2013), where inhomogeneous Poisson processes are considered. Theorem 2.1 differs in two aspects from their Theorem 1. Firstly, we do not confine ourselves to inhomogeneous Poisson processes. Secondly and more importantly, our conditions are different: we do not assume that  $\lambda_0$  is bounded below away from zero and we do not need to bound from below the prior mass in neighborhoods of  $\lambda_0$  for the sup-norm, rather the prior mass in neighborhoods of  $\lambda_0$  for the Hellinger distance, as in Theorem 2.2 of Ghosal et al. (2000). In Theorem 2.1, our aim is to propose conditions to assess posterior concentration rates for intensity functions resembling those used in the density model obtained by parameterizing  $\lambda$  as  $\lambda = M_\lambda \times \bar{\lambda}$ , with  $\bar{\lambda}$  a probability density on  $\Omega$ .

**Remark 2.1.** If  $\bar{\lambda} \in \bar{B}_{2,n}(\bar{\lambda}_0; v_n, H)$  then, for every integer  $j > 2$ ,  $E_j(\bar{\lambda}_0; \bar{\lambda}) \leq H^{j-2} v_n^2 (\log n)^{j-2}$  so that, using Proposition 4.1, if we replace  $\bar{B}_{k,n}(\bar{\lambda}_0; v_n, H)$  with  $\bar{B}_{2,n}(\bar{\lambda}_0; v_n, H)$  in the assumptions of Theorem 2.1, we obtain the same type of conclusion: for any  $k \geq 2$  such that condition (2.1) is satisfied, we have

$$\mathbb{E}_{\lambda_0}^{(n)}[\pi(\lambda : \|\lambda - \lambda_0\|_1 > J_1 v_n \mid N)] = O((nv_n^2)^{-k/2} (\log n)^{k(k_{[2]}-2)/2}),$$

with an extra  $(\log n)$ -term on the right-hand side of the above equality.

**Remark 2.2.** Condition (2.1) is satisfied for the above considered examples: it is verified for inhomogeneous Poisson processes since  $Y_t = n$  for every  $t$ . For the censoring model,  $Y_t = \sum_{i=1}^n \mathbf{1}_{Z_i \geq t}$ . For every  $i = 1, \dots, n$ , we set  $V_i = \mathbf{1}_{Z_i \geq t} - \mathbb{P}(Z_1 \geq t)$ . Then, for  $k \geq 2$ ,

$$\begin{aligned} \mathbb{E}_{\lambda_0}^{(n)} \left[ \left( \int_{\Omega} [Y_t - \mu_n(t)]^2 dt \right)^k \right] &= \mathbb{E}_{\lambda_0}^{(n)} \left[ \left( \int_0^T \left( \sum_{i=1}^n V_i \right)^2 dt \right)^k \right] \\ &\lesssim \int_0^T \mathbb{E}_{\lambda_0}^{(n)} \left[ \left( \sum_{i=1}^n V_i \right)^{2k} \right] dt \\ &\lesssim \int_0^T \left( \sum_{i=1}^n \mathbb{E}_{\lambda_0}^{(n)} [V_i^{2k}] + \left( \sum_{i=1}^n \mathbb{E}_{\lambda_0}^{(n)} [V_i^2] \right)^k \right) dt \lesssim n^k \end{aligned}$$

by Hölder and Rosenthal inequalities (see, for instance, Theorem C.2 of Härdle et al. (1998)). Under mild conditions, similar computations can be performed for finite state Markov processes.

Conditions of Theorem 2.1 are very similar to those considered for density estimation in the case of i.i.d. observations. In particular,

$$\bar{B}_n = \left\{ \bar{\lambda} : h^2(\bar{\lambda}_0, \bar{\lambda}) \left\| \frac{\bar{\lambda}_0}{\bar{\lambda}} \right\|_{\infty} \leq v_n^2, \left\| \frac{\bar{\lambda}_0}{\bar{\lambda}} \right\|_{\infty} \leq n^H, \|\bar{\lambda}\|_{\infty} \leq H \right\}$$

is included in  $\bar{B}_{k,n}(\bar{\lambda}_0; v_n(\log n)^{1/2}, H)$  as a consequence of Theorem 5.1 of Wong and Shen (1995). Apart from the mild constraints  $\|\bar{\lambda}_0/\bar{\lambda}\|_{\infty} \leq n^H$  and  $\|\bar{\lambda}\|_{\infty} \leq H$ , the set  $\bar{B}_n$  is the same as the one considered in Theorem 2.2 of Ghosal et al. (2000). The other conditions are essentially those of Theorem 2.1 in Ghosal et al. (2000).

### 3 Illustrations with different families of priors

As discussed in Section 2, the conditions of Theorem 2.1 to derive posterior contraction rates are very similar to those considered in the literature for density estimation so that existing results involving different families of prior distributions can be adapted to Aalen multiplicative intensity models. Some applications are presented below.

#### 3.1 Monotone non-increasing intensity functions

In this section, we deal with estimation of monotone non-increasing intensity functions, which is equivalent to considering monotone non-increasing density functions  $\bar{\lambda}$  in the above described parametrization. To construct a prior on the set of monotone non-increasing densities over  $[0, T]$ , we use their representation as mixtures of uniform

densities as in Williamson (1956) and consider a Dirichlet process as a prior on the mixing distribution:

$$\bar{\lambda}(\cdot) = \int_0^\infty \frac{\mathbf{1}_{(0,\theta)}(\cdot)}{\theta} dP(\theta), \quad P \mid A, G \sim \text{DP}(AG), \quad (3.1)$$

where  $G$  is a distribution on  $[0, T]$  having density  $g$  with respect to Lebesgue measure. This prior has been studied by Salomond (2013) for estimating monotone non-increasing densities. Here, we extend his results to the case of monotone non-increasing intensity functions of Aalen processes. We consider the same assumption on  $G$  as in Salomond (2013): there exist  $a_1, a_2 > 0$  such that

$$\theta^{a_1} \lesssim g(\theta) \lesssim \theta^{a_2} \quad \text{for all } \theta \text{ in a neighbourhood of } 0. \quad (3.2)$$

The following result holds.

**Corollary 3.1.** *Assume that the counting process  $N$  verifies conditions (1.2) and (1.3) and that inequality (2.1) is satisfied for some  $k \geq 2$ . Consider a prior  $\pi_1$  on  $\bar{\lambda}$  satisfying conditions (3.1) and (3.2) and a prior  $\pi_M$  on  $M_\lambda$  that is absolutely continuous with respect to Lebesgue measure with positive and continuous density on  $\mathbb{R}_+$ . Suppose that  $\lambda_0$  is monotone non-increasing and bounded on  $\mathbb{R}_+$ . Let  $\bar{\epsilon}_n = (n/\log n)^{-1/3}$ . Then, there exists a constant  $J_1 > 0$  such that*

$$\mathbb{E}_{\lambda_0}^{(n)}[\pi(\lambda : \|\lambda - \lambda_0\|_1 > J_1 \bar{\epsilon}_n \mid N)] = O((n\bar{\epsilon}_n^2)^{-k/2} (\log n)^{k(k_{[2]}-2)/2}).$$

The proof is reported in Section 4.

### 3.2 Log-spline and log-linear priors on $\lambda$

For simplicity of presentation, we set  $T = 1$ . We consider a log-spline prior of order  $q$  as in Section 4 of Ghosal et al. (2000). In other words,  $\bar{\lambda}$  is parameterized as

$$\log \bar{\lambda}_\theta(\cdot) = \theta^t \underline{B}_J(\cdot) - c(\theta), \quad \text{with } \exp(c(\theta)) = \int_0^1 e^{\theta^t \underline{B}_J(x)} dx,$$

where  $\underline{B}_J = (B_1, \dots, B_J)$  is the  $q$ -th order  $B$ -spline defined in de Boor (1978) associated with  $K$  fixed knots, so that  $J = K + q - 1$ , see Ghosal et al. (2000) for more details. Consider a prior on  $\theta$  in the form  $J = J_n = \lfloor n^{1/(2\alpha+1)} \rfloor$ ,  $\alpha \in [1/2, q]$  and, conditionally on  $J$ , the prior is absolutely continuous with respect to Lebesgue measure on  $[-M, M]^J$  with density bounded from below and above by  $c^J$  and  $C^J$ , respectively. Consider an absolutely continuous prior with positive and continuous density on  $\mathbb{R}_+$  on  $M_\lambda$ . We then have the following posterior concentration result.

**Corollary 3.2.** *For the above prior, if  $\|\log \lambda_0\|_\infty < \infty$  and  $\lambda_0$  is Hölder with regularity  $\alpha \in [1/2, q]$ , then under condition (2.1), there exists a constant  $J_1 > 0$  so that*

$$\mathbb{E}_{\lambda_0}^{(n)}[\pi(\lambda : \|\lambda - \lambda_0\|_1 > J_1 n^{-\alpha/(2\alpha+1)} \mid N)] = O(n^{-k/(4\alpha+2)} (\log n)^{k(k_{[2]}-2)/2}).$$

*Proof.* Set  $\epsilon_n = n^{-\alpha/(2\alpha+1)}$ . Using Lemma 4.1, there exists  $\theta_0 \in \mathbb{R}^J$  such that  $h(\bar{\lambda}_{\theta_0}, \bar{\lambda}_0) \lesssim \|\log \bar{\lambda}_{\theta_0} - \log \bar{\lambda}_0\|_\infty \lesssim J^{-\alpha}$ , which combined with Lemma 4.4 leads to

$$\pi_1(\bar{B}_{k,n}(\bar{\lambda}_0; \epsilon_n, H) \geq e^{-C_1 n \epsilon_n^2}.$$

Lemma 4.5 together with Theorem 4.5 of Ghosal et al. (2000) controls the entropy of  $\bar{S}_{n,j}$  and its prior mass for  $j$  larger than some fixed constant  $J_0$ .  $\square$

With such families of priors, it is more interesting to work with non-normalized  $\lambda_\theta$ . We can write

$$\lambda_{A,\theta}(\cdot) = A \exp(\theta^t \underline{B}_J(\cdot)), \quad A > 0,$$

so that a prior on  $\lambda$  is defined as a prior on  $A$ , say  $\pi_A$  absolutely continuous with respect to Lebesgue measure having positive and continuous density and the same type of prior prior on  $\theta$  as above. The same result then holds. It is not a direct consequence of Theorem 2.1, since  $M_{\lambda_{A,\theta}} = A \exp(c(\theta))$  is not *a priori* independent of  $\bar{\lambda}_{A,\theta}$ . However, introducing  $A$  allows to adapt Theorem 2.1 to this case. The practical advantage of the latter representation is that it avoids computing the normalizing constant  $c(\theta)$ .

In a similar manner, we can replace spline basis with other orthonormal bases, as considered in Rivoirard and Rousseau (2012), leading to the same posterior concentration rates as in density estimation. More precisely, consider intensities parameterized as

$$\bar{\lambda}_\theta(\cdot) = e^{\sum_{j=1}^J \theta_j \phi_j(\cdot) - c(\theta)}, \quad e^{c(\theta)} = \int_{\mathbb{R}^J} e^{\sum_{j=1}^J \theta_j \phi_j(x)} dx,$$

where  $(\phi_j)_{j=1}^\infty$  is an orthonormal basis of  $L_2([0, 1])$ , with  $\phi_1 = 1$ . Write  $\eta = (A, \theta)$ , with  $A > 0$ , and

$$\lambda_\eta(\cdot) = A e^{\sum_{j=1}^J \theta_j \phi_j(\cdot)} = A e^{c(\theta)} \bar{\lambda}_\theta(\cdot).$$

Let  $A \sim \pi_A$  and consider the same family of priors as in Rivoirard and Rousseau (2012):

$$J \sim \pi_J,$$

$$j^\beta \theta_j / \tau_0 \stackrel{\text{ind}}{\sim} g, \quad j \leq J, \quad \text{and} \quad \theta_j = 0, \quad \forall j > J,$$

where  $g$  is a positive and continuous density on  $\mathbb{R}$  and there exist  $s \geq 0$  and  $p > 0$  such that

$$\log \pi_J(J) \asymp -J(\log J)^s, \quad \log g(x) \asymp -|x|^p, \quad s = 0, 1,$$

when  $J$  and  $|x|$  are large. Rivoirard and Rousseau (2012) prove that this prior leads to minimax adaptive posterior concentration rates over collections of positive and Hölder classes of densities in the density model. Their proof easily extends to prove assumptions (i) and (ii) of Theorem 2.1.

**Corollary 3.3.** *Consider the above described prior on an intensity function  $\lambda$  on  $[0, 1]$ . Assume that  $\lambda_0$  is positive and belongs to a Sobolev class with smoothness  $\alpha > 1/2$ . Under condition (2.1), if  $\beta < 1/2 + \alpha$ , there exists a constant  $J_1 > 0$  so that*

$$\begin{aligned} \mathbb{E}_{\lambda_0}^{(n)}[\pi(\lambda : \|\lambda - \lambda_0\|_1 > J_1(n/\log n)^{-\alpha/(2\alpha+1)}(\log n)^{(1-s)/2} \mid N)] \\ = O(n^{-k/(4\alpha+2)}(\log n)^{k(k_{[2]}-2)/2}). \end{aligned}$$



Note that the constraint  $\beta < \alpha + 1/2$  is satisfied for all  $\alpha > 1/2$  as soon as  $\beta < 1$  and, as in Rivoirard and Rousseau (2012), the prior leads to adaptive minimax posterior concentration rates over collections of Sobolev balls.

## 4 Proofs

To prove Theorem 2.1, we use the following intermediate results whose proofs are postponed to Section 5. The first one controls the Kullback-Leibler divergence and absolute moments of  $\ell_n(\lambda_0) - \ell_n(\lambda)$ , where  $\ell_n(\lambda)$  is the log-likelihood for Aalen processes evaluated at  $\lambda$ , whose expression is given by

$$\ell_n(\lambda) = \int_0^T \log(\lambda(t)) dN_t - \int_0^T \lambda(t) Y_t dt,$$

see Andersen et al. (1993).

**Proposition 4.1.** *Let  $v_n$  be a positive sequence such that  $v_n \rightarrow 0$  and  $nv_n^2 \rightarrow \infty$ . For any  $k \geq 2$  and  $H > 0$ , define the set*

$$B_{k,n}(\lambda_0; v_n, H) = \{\lambda : \bar{\lambda} \in \bar{B}_{k,n}(\bar{\lambda}_0; v_n, H), |M_\lambda - M_{\lambda_0}| \leq v_n\}.$$

*Under assumptions (1.2) and (2.1), for all  $\lambda \in B_{k,n}(\lambda_0; v_n, H)$ , we have*

$$\text{KL}(\lambda_0; \lambda) \leq \kappa_0 nv_n^2 \quad \text{and} \quad V_k(\lambda_0; \lambda) \leq \kappa (nv_n^2)^{k/2},$$

*where  $\kappa_0, \kappa$  depend only on  $k, C_{1k}, H, \lambda_0, m_1$  and  $m_2$ . An expression of  $\kappa_0$  is given in (2.2).*

The second result establishes the existence of tests that are used to control the numerator of posterior distributions. We use that, under assumption (1.2), on the set  $\Gamma_n$ ,

$$\forall t \in \Omega, \quad (1 - \alpha)\tilde{\mu}_n(t) \leq \frac{Y_t}{n} \leq (1 + \alpha)\tilde{\mu}_n(t). \quad (4.1)$$

**Proposition 4.2.** *Assume that conditions (i) and (ii) of Theorem 2.1 are satisfied. For any  $j \in \mathbb{N}$ , define*

$$S_{n,j}(v_n) = \{\lambda : \bar{\lambda} \in \mathcal{F}_n \text{ and } jv_n < \|\lambda - \lambda_0\|_1 \leq (j+1)v_n\}.$$

*Then, under assumption (1.2), there are constants  $J_0, \rho, c > 0$  such that, for every integer  $j \geq J_0$ , there exists a test  $\phi_{n,j}$  so that, for a positive constant  $C$ ,*

$$\mathbb{E}_{\lambda_0}^{(n)}[\mathbf{1}_{\Gamma_n} \phi_{n,j}] \leq Ce^{-cnj^2 v_n^2}, \quad \sup_{\lambda \in S_{n,j}(v_n)} \mathbb{E}_\lambda[\mathbf{1}_{\Gamma_n} (1 - \phi_{n,j})] \leq Ce^{-cnj^2 v_n^2}, \quad J_0 \leq j \leq \frac{\rho}{v_n},$$

*and*

$$\mathbb{E}_{\lambda_0}^{(n)}[\mathbf{1}_{\Gamma_n} \phi_{n,j}] \leq Ce^{-cnjv_n}, \quad \sup_{\lambda \in S_{n,j}(v_n)} \mathbb{E}_\lambda[\mathbf{1}_{\Gamma_n} (1 - \phi_{n,j})] \leq Ce^{-cnjv_n}, \quad j > \frac{\rho}{v_n}.$$

In what follows, the symbols “ $\lesssim$ ” and “ $\gtrsim$ ” are used to denote inequalities valid up to constants that are universal or fixed throughout.

*Proof of Theorem 2.1.* Given Proposition 4.1 and Proposition 4.2, the proof of Theorem 2.1 is similar to that of Theorem 1 in Ghosal and van der Vaart (2007). Let  $U_n = \{\lambda : \|\lambda - \lambda_0\|_1 > J_1 v_n\}$ . Write

$$\pi(U_n \mid N) = \frac{\int_{U_n} e^{\ell_n(\lambda) - \ell_n(\lambda_0)} d\pi(\lambda)}{\int_{\mathcal{F}} e^{\ell_n(\lambda) - \ell_n(\lambda_0)} d\pi(\lambda)} = \frac{N_n}{D_n}.$$

We have

$$\begin{aligned} \mathbb{P}_{\lambda_0}^{(n)} \left( D_n \leq e^{-(\kappa_0+1)nv_n^2} \pi_1(\bar{B}_{k,n}(\bar{\lambda}_0; v_n, H)) \right) \\ \leq \mathbb{P}_{\lambda_0}^{(n)} \left( \int_{B_{k,n}(\lambda_0; v_n, H)} \frac{\exp\{\ell_n(\lambda) - \ell_n(\lambda_0)\}}{\pi(B_{k,n}(\lambda_0; v_n, H))} d\pi(\lambda) \right. \\ \left. \leq -(\kappa_0 + 1)nv_n^2 + \log \left( \frac{\pi_1(\bar{B}_{k,n}(\bar{\lambda}_0; v_n, H))}{\pi(B_{k,n}(\lambda_0; v_n, H))} \right) \right). \end{aligned}$$

By the assumption on the positivity and continuity of the Lebesgue density of the prior  $\pi_M$  and the requirement that  $v_n^2 \geq (n/\log n)^{-1}$ ,

$$\pi(B_{k,n}(\lambda_0; v_n, H)) \gtrsim \pi_1(\bar{B}_{k,n}(\bar{\lambda}_0; v_n, H))v_n \gtrsim \pi_1(\bar{B}_{k,n}(\bar{\lambda}_0; v_n, H))e^{-nv_n^2/2},$$

so that, using Proposition 4.1 and Markov's inequality,

$$\mathbb{P}_{\lambda_0}^{(n)} \left( D_n \leq e^{-(\kappa_0+1)nv_n^2} \pi_1(\bar{B}_{k,n}(\bar{\lambda}_0; v_n, H)) \right) \lesssim (nv_n^2)^{-k/2}.$$

Note that inequality (5.6) implies that  $\pi(S_{n,j}(v_n)) \leq \pi_1(\bar{S}_{n,j})$ . Using tests  $\phi_{n,j}$  of Proposition 4.2, mimicking the proof of Theorem 1 of Ghosal and van der Vaart (2007), we have that for  $J_1 \geq J_0$ ,

$$\begin{aligned} \mathbb{E}_{\lambda_0}^{(n)} [\mathbf{1}_{\Gamma_n} \pi(\lambda : \|\lambda - \lambda_0\|_1 > J_1 v_n \mid N)] \\ \leq \sum_{j \geq J_1} \mathbb{E}_{\lambda_0}^{(n)} [\mathbf{1}_{\Gamma_n} \phi_{n,j}] + \sum_{j=\lceil J_1 \rceil}^{\lfloor \rho/v_n \rfloor} e^{(\kappa_0+1)nv_n^2} \frac{\pi_1(\bar{S}_{n,j})e^{-cnj^2v_n^2}}{\pi_1(\bar{B}_{k,n}(\bar{\lambda}_0; v_n, H))} \\ + \sum_{j > \rho/v_n} \frac{e^{(\kappa_0+1)nv_n^2} \pi_1(\bar{S}_{n,j})e^{-cnjv_n}}{\pi_1(\bar{B}_{k,n}(\bar{\lambda}_0; v_n, H))} + \frac{e^{(\kappa_0+1)nv_n^2} \pi_1(\mathcal{F}_n^c)}{\pi_1(\bar{B}_{k,n}(\bar{\lambda}_0; v_n, H))} \\ + \mathbb{P}_{\lambda_0}^{(n)}(D_n \leq e^{-(\kappa_0+1)nv_n^2} \pi_1(\bar{B}_{k,n}(\bar{\lambda}_0; v_n, H))) \\ \lesssim (nv_n^2)^{-k/2}, \end{aligned}$$

which proves the result since  $\mathbb{P}_{\lambda_0}^{(n)}(\Gamma_n^c) = o(1)$ .  $\square$

*Proof of Corollary 3.1.* Without loss of generality, we can assume that  $\Omega = [0, T]$ . At several places, using (1.1) and (4.1), we have that, under  $\mathbb{P}_\lambda^{(n)}(\cdot \mid \Gamma_n)$ , for any interval  $I$ , the number of points of  $N$  falling in  $I$  is controlled by the number of points of a Poisson process with intensity  $n(1 + \alpha)m_2\lambda$  falling in  $I$ . Recall that  $\bar{\epsilon}_n = (n/\log n)^{-1/3}$ . For  $\kappa_0$  as in (2.2), we control  $\mathbb{P}_{\lambda_0}^{(n)}(\ell_n(\lambda) - \ell_n(\lambda_0) \leq -(\kappa_0 + 1)n\bar{\epsilon}_n^2)$ . We follow most of the computations of Salomond (2013). Let  $e_n = (n\bar{\epsilon}_n^2)^{-k/2}$ ,

$$\bar{\lambda}_{0n}(t) = \frac{\lambda_0(t)\mathbf{1}_{t \leq \theta_n}}{\int_0^{\theta_n} \lambda_0(u)du}, \quad \text{with } \theta_n = \inf \left\{ \theta : \int_0^\theta \bar{\lambda}_0(t)dt \geq 1 - \frac{e_n}{n} \right\},$$

and  $\lambda_{0n} = M_{\lambda_0} \bar{\lambda}_{0n}$ . Define the event  $A_n = \{X \in N : X \leq \theta_n\}$ . We make use of the following result. Let  $\tilde{N}$  be a Poisson process with intensity  $n(1 + \alpha)m_2\lambda_0$ . If  $\tilde{N}(T) = k$ , denote by  $\tilde{N} = \{X_1, \dots, X_k\}$ . Conditionally on  $\tilde{N}(T) = k$ , the random variables  $X_1, \dots, X_k$  are i.i.d. with density  $\bar{\lambda}_0$ . So,

$$\begin{aligned} \mathbb{P}_{\lambda_0}^{(n)}(A_n^c \mid \Gamma_n) &\leq \sum_{k=1}^{\infty} \mathbb{P}_{\lambda_0}^{(n)}(\exists X_i > \theta_n \mid \tilde{N}(T) = k) \mathbb{P}_{\lambda_0}^{(n)}(\tilde{N}(T) = k) \\ &\leq \sum_{k=1}^{\infty} \left(1 - \left(1 - \frac{e_n}{n}\right)^k\right) \mathbb{P}_{\lambda_0}^{(n)}(\tilde{N}(T) = k) \\ &= O\left(\frac{e_n}{n} \mathbb{E}_{\lambda_0}^{(n)}[\tilde{N}(T)]\right) = O(e_n) = O((n\bar{\epsilon}_n^2)^{-k/2}). \end{aligned}$$

Now,

$$\begin{aligned} \mathbb{P}_{\lambda_0}^{(n)}(\ell_n(\lambda) - \ell_n(\lambda_0) \leq -(\kappa_0 + 2)n\bar{\epsilon}_n^2 \mid \Gamma_n) \\ \leq \mathbb{P}_{\lambda_0}^{(n)}(\ell_n(\lambda) - \ell_n(\lambda_0) \leq -(\kappa_0 + 2)n\bar{\epsilon}_n^2 \mid A_n, \Gamma_n) + \mathbb{P}_{\lambda_0}^{(n)}(A_n^c \mid \Gamma_n). \end{aligned}$$

We now deal with the first term on the right-hand side. On  $\Gamma_n \cap A_n$ ,

$$\begin{aligned} \ell_n(\lambda_0) &= \ell_n(\lambda_{0n}) + \int_0^{\theta_n} \log\left(\frac{\lambda_0(t)}{\lambda_{0n}(t)}\right) dN_t - \int_0^T [\lambda_0(t) - \lambda_{0n}(t)] Y_t dt \\ &= \ell_n(\lambda_{0n}) + N(T) \log\left(\int_0^{\theta_n} \bar{\lambda}_0(t) dt\right) - M_{\lambda_0} \int_0^T \bar{\lambda}_0(t) Y_t dt + M_{\lambda_0} \frac{\int_0^{\theta_n} \bar{\lambda}_0(t) Y_t dt}{\int_0^{\theta_n} \bar{\lambda}_0(t) dt} \\ &\leq \ell_n(\lambda_{0n}) + M_{\lambda_0} \frac{\int_{\theta_n}^T \bar{\lambda}_0(t) dt \int_0^{\theta_n} \bar{\lambda}_0(t) Y_t dt}{\int_0^{\theta_n} \bar{\lambda}_0(t) dt} - M_{\lambda_0} \int_{\theta_n}^T \bar{\lambda}_0(t) Y_t dt \\ &\leq \ell_n(\lambda_{0n}) + M_{\lambda_0} \frac{e_n(1 + \alpha)m_2}{1 - e_n/n}. \end{aligned}$$

So, for every  $\lambda$  and any  $n$  large enough,

$$\begin{aligned} \mathbb{P}_{\lambda_0}^{(n)}(\ell_n(\lambda) - \ell_n(\lambda_0) \leq -(\kappa_0 + 2)n\bar{\epsilon}_n^2 \mid A_n, \Gamma_n) \\ \leq \mathbb{P}_{\lambda_0}^{(n)}(\ell_n(\lambda) - \ell_n(\lambda_{0n}) \leq -(\kappa_0 + 1)n\bar{\epsilon}_n^2 \mid A_n, \Gamma_n) \\ = \mathbb{P}_{\lambda_{0n}}^{(n)}(\ell_n(\lambda) - \ell_n(\lambda_{0n}) \leq -(\kappa_0 + 1)n\bar{\epsilon}_n^2 \mid \Gamma_n) \end{aligned}$$

because  $\mathbb{P}_{\lambda_0}^{(n)}(\cdot | A_n) = \mathbb{P}_{\lambda_{0n}}^{(n)}(\cdot)$ . Let  $H > 0$  be fixed. For all  $\lambda \in B_{k,n}(\lambda_{0n}; \bar{\epsilon}_n, H)$ , using Proposition 4.1, we obtain

$$\mathbb{P}_{\lambda_{0n}}^{(n)}(\ell_n(\lambda) - \ell_n(\lambda_{0n}) \leq -(\kappa_0 + 1)n\bar{\epsilon}_n^2 | \Gamma_n) = O((n\bar{\epsilon}_n^2)^{-k/2}).$$

Mimicking the proof of Lemma 8 in Salomond (2013), we have that, for some constant  $C_k > 0$ ,

$$\pi_1(\bar{B}_{k,n}(\bar{\lambda}_{0n}; \bar{\epsilon}_n, H)) \geq e^{-C_k n \bar{\epsilon}_n^2} \quad \text{when } n \text{ is large enough,}$$

so that the first part of condition (ii) of Theorem 2.1 is verified. As in Salomond (2013), we set  $\mathcal{F}_n = \{\bar{\lambda} : \bar{\lambda}(0) \leq M_n\}$ , with  $M_n = \exp(c_1 n \bar{\epsilon}_n^2)$  and  $c_1$  a positive constant. From Lemma 9 of Salomond (2013), there exists  $a > 0$  such that  $\pi_1(\mathcal{F}_n^c) \leq e^{-c_1(a+1)n\bar{\epsilon}_n^2}$  for  $n$  large enough, and the first part of condition (i) is satisfied. It is known from Groeneboom (1985) that the  $\epsilon$ -entropy of  $\mathcal{F}_n$  is of order  $(\log M_n)/\epsilon$ , that is  $o(n)$  for all  $\epsilon > 0$  and the second part of (i) holds. The second part of (ii) is a consequence of equation (22) of Salomond (2013).  $\square$

## 5 Auxiliary results

This section reports the proofs of Proposition 4.1 and Proposition 4.2 that have been stated in Section 4. Proofs of intermediate results are deferred to Section 6.

We use the fact that for any pair of densities  $f$  and  $g$ ,  $\|f - g\|_1 \leq 2h(f, g)$ .

*Proof of Proposition 4.1.* Recall that the log-likelihood evaluated at  $\lambda$  is given by  $\ell_n(\lambda) = \int_0^T \log(\lambda(t)) dN_t - \int_0^T \lambda(t) Y_t dt$ . Since on  $\Omega^c$ ,  $N$  is empty and  $Y_t \equiv 0$  almost surely, we can assume, without loss of generality, that  $\Omega = [0, T]$ . Define

$$M_n(\lambda) = \int_0^T \lambda(t) \mu_n(t) dt, \quad M_n(\lambda_0) = \int_0^T \lambda_0(t) \mu_n(t) dt,$$

and

$$\bar{\lambda}_n(\cdot) = \frac{\lambda(\cdot) \mu_n(\cdot)}{M_n(\lambda)} = \frac{\bar{\lambda}(\cdot) \tilde{\mu}_n(\cdot)}{\int_0^T \bar{\lambda}(t) \tilde{\mu}_n(t) dt}, \quad \bar{\lambda}_{0,n}(\cdot) = \frac{\lambda_0(\cdot) \mu_n(\cdot)}{M_n(\lambda_0)} = \frac{\bar{\lambda}_0(\cdot) \tilde{\mu}_n(\cdot)}{\int_0^T \bar{\lambda}_0(t) \tilde{\mu}_n(t) dt}.$$

By straightforward computations,

$$\begin{aligned} \text{KL}(\lambda_0; \lambda) &= \mathbb{E}_{\lambda_0}^{(n)}[\ell_n(\lambda_0) - \ell_n(\lambda)] \\ &= M_n(\lambda_0) \left[ \text{KL}(\bar{\lambda}_{0,n}; \bar{\lambda}_n) + \frac{M_n(\lambda)}{M_n(\lambda_0)} - 1 - \log \left( \frac{M_n(\lambda)}{M_n(\lambda_0)} \right) \right] \\ &= M_n(\lambda_0) \left[ \text{KL}(\bar{\lambda}_{0,n}; \bar{\lambda}_n) + \phi \left( \frac{M_n(\lambda)}{M_n(\lambda_0)} \right) \right] \\ &\leq nm_2 M_{\lambda_0} \left[ \text{KL}(\bar{\lambda}_{0,n}; \bar{\lambda}_n) + \phi \left( \frac{M_n(\lambda)}{M_n(\lambda_0)} \right) \right], \end{aligned} \tag{5.1}$$

where  $\phi(x) = x - 1 - \log x$  and

$$\text{KL}(\bar{\lambda}_{0,n}; \bar{\lambda}_n) = \int_0^T \log \left( \frac{\bar{\lambda}_{0,n}(t)}{\bar{\lambda}_n(t)} \right) \bar{\lambda}_{0,n}(t) dt.$$

We control  $\text{KL}(\bar{\lambda}_{0,n}; \bar{\lambda}_n)$  for  $\lambda \in B_{k,n}(\lambda_0; v_n, H)$ . By using Lemma 8.2 of Ghosal et al. (2000), we have

$$\begin{aligned} \text{KL}(\bar{\lambda}_{0,n}; \bar{\lambda}_n) &\leq 2h^2(\bar{\lambda}_{0,n}, \bar{\lambda}_n) \left( 1 + \log \left\| \frac{\bar{\lambda}_{0,n}}{\bar{\lambda}_n} \right\|_\infty \right) \\ &\leq 2h^2(\bar{\lambda}_{0,n}, \bar{\lambda}_n) \left[ 1 + \log \left( \frac{m_2}{m_1} \right) + \log \left\| \frac{\bar{\lambda}_0}{\bar{\lambda}} \right\|_\infty \right] \\ &\leq 2 \left[ 1 + \log \left( \frac{m_2}{m_1} \right) \right] h^2(\bar{\lambda}_{0,n}, \bar{\lambda}_n) \left( 1 + \log \left\| \frac{\bar{\lambda}_0}{\bar{\lambda}} \right\|_\infty \right) \end{aligned} \quad (5.2)$$

because  $1 + \log(m_2/m_1) \geq 1$ . We now deal with  $h^2(\bar{\lambda}_{0,n}, \bar{\lambda}_n)$ . We have

$$\begin{aligned} h^2(\bar{\lambda}_{0,n}, \bar{\lambda}_n) &= \int_0^T \left( \sqrt{\bar{\lambda}_{0,n}(t)} - \sqrt{\bar{\lambda}_n(t)} \right)^2 dt \\ &= \int_0^T \left( \sqrt{\frac{\bar{\lambda}_0(t)\tilde{\mu}_n(t)}{\int_0^T \bar{\lambda}_0(u)\tilde{\mu}_n(u)du}} - \sqrt{\frac{\bar{\lambda}(t)\tilde{\mu}_n(t)}{\int_0^T \bar{\lambda}(u)\tilde{\mu}_n(u)du}} \right)^2 dt \\ &\leq 2m_2 \int_0^T \left( \sqrt{\frac{\bar{\lambda}_0(t)}{\int_0^T \bar{\lambda}_0(u)\tilde{\mu}_n(u)du}} - \sqrt{\frac{\bar{\lambda}_0(t)}{\int_0^T \bar{\lambda}(u)\tilde{\mu}_n(u)du}} \right)^2 dt \\ &\quad + 2m_2 \int_0^T \left( \sqrt{\frac{\bar{\lambda}_0(t)}{\int_0^T \bar{\lambda}(u)\tilde{\mu}_n(u)du}} - \sqrt{\frac{\bar{\lambda}(t)}{\int_0^T \bar{\lambda}(u)\tilde{\mu}_n(u)du}} \right)^2 dt \\ &\leq 2m_2 U_n + \frac{2m_2}{m_1} h^2(\bar{\lambda}_0, \bar{\lambda}), \end{aligned}$$

with

$$U_n = \left( \sqrt{\frac{1}{\int_0^T \bar{\lambda}_0(t)\tilde{\mu}_n(t)dt}} - \sqrt{\frac{1}{\int_0^T \bar{\lambda}(t)\tilde{\mu}_n(t)dt}} \right)^2.$$

We denote by

$$\tilde{\epsilon}_n := \frac{1}{\int_0^T \bar{\lambda}_0(u)\tilde{\mu}_n(u)du} \int_0^T [\bar{\lambda}(t) - \bar{\lambda}_0(t)]\tilde{\mu}_n(t)dt,$$

so that

$$|\tilde{\epsilon}_n| \leq \frac{1}{m_1} \int_0^T |\bar{\lambda}(t) - \bar{\lambda}_0(t)|\tilde{\mu}_n(t)dt \leq \frac{2m_2}{m_1} h(\bar{\lambda}_0, \bar{\lambda}).$$

Then,

$$U_n = \frac{1}{\int_0^T \bar{\lambda}_0(t)\tilde{\mu}_n(t)dt} \left( 1 - \frac{1}{\sqrt{1 + \tilde{\epsilon}_n}} \right)^2 \leq \frac{\tilde{\epsilon}_n^2}{4m_1} \leq \frac{m_2^2}{m_1^3} h^2(\bar{\lambda}_0, \bar{\lambda}).$$

Finally,

$$h^2(\bar{\lambda}_{0,n}, \bar{\lambda}_n) \leq \frac{2m_2}{m_1} \left( \frac{m_2^2}{m_1^2} + 1 \right) h^2(\bar{\lambda}_0, \bar{\lambda}). \quad (5.3)$$

It remains to bound  $\phi(M_n(\lambda)/M_n(\lambda_0))$ . We have

$$\begin{aligned} |M_n(\lambda_0) - M_n(\lambda)| &\leq \int_0^T |\lambda(t) - \lambda_0(t)| \mu_n(t) dt \\ &\leq nm_2 \int_0^T |\lambda(t) - \lambda_0(t)| dt \\ &\leq \frac{m_2}{m_1 M_{\lambda_0}} M_n(\lambda_0) [M_{\lambda_0} \|\bar{\lambda} - \bar{\lambda}_0\|_1 + |M_\lambda - M_{\lambda_0}|] \\ &\leq \frac{m_2}{m_1 M_{\lambda_0}} M_n(\lambda_0) [2M_{\lambda_0} h(\bar{\lambda}, \bar{\lambda}_0) + |M_\lambda - M_{\lambda_0}|] \\ &\leq \frac{m_2}{m_1 M_{\lambda_0}} M_n(\lambda_0) (2M_{\lambda_0} + 1) v_n. \end{aligned}$$

Since  $\phi(u+1) \leq u^2$  if  $|u| \leq 1/2$ , we have

$$\phi\left(\frac{M_n(\lambda)}{M_n(\lambda_0)}\right) \leq \frac{m_2^2}{m_1^2 M_{\lambda_0}^2} (2M_{\lambda_0} + 1)^2 v_n^2 \quad \text{for } n \text{ large enough.} \quad (5.4)$$

Combining (5.1), (5.2), (5.3) and (5.4), we have  $\text{KL}(\lambda_0; \lambda) \leq \kappa_0 n v_n^2$  for  $n$  large enough, with  $\kappa_0$  as in (2.2). We now deal with

$$V_{2k}(\lambda_0; \lambda) = \mathbb{E}_{\lambda_0}^{(n)}[\ell_n(\lambda_0) - \ell_n(\lambda) - \mathbb{E}_{\lambda_0}^{(n)}[\ell_n(\lambda_0) - \ell_n(\lambda)]^{2k}], \quad k \geq 1.$$

We begin by considering the case  $k > 1$ . In the sequel, we denote by  $C$  a constant that may change from line to line. Straightforward computations lead to

$$\begin{aligned} V_{2k}(\lambda_0; \lambda) &= \mathbb{E}_{\lambda_0}^{(n)} \left[ \left| - \int_0^T \left[ \lambda_0(t) - \lambda(t) - \lambda_0(t) \log \left( \frac{\lambda_0(t)}{\lambda(t)} \right) \right] [Y_t - \mu_n(t)] dt \right. \right. \\ &\quad \left. \left. + \int_0^T \log \left( \frac{\lambda_0(t)}{\lambda(t)} \right) [dN_t - Y_t \lambda_0(t) dt] \right|^{2k} \right] \\ &\leq 2^{2k-1} (A_{2k} + B_{2k}), \end{aligned}$$

with

$$B_{2k} := \mathbb{E}_{\lambda_0}^{(n)} \left[ \left| \int_0^T \log \left( \frac{\lambda_0(t)}{\lambda(t)} \right) [dN_t - Y_t \lambda_0(t) dt] \right|^{2k} \right]$$

and, by (2.1),

$$\begin{aligned}
A_{2k} &:= \mathbb{E}_{\lambda_0}^{(n)} \left[ \left| \int_0^T \left[ \lambda_0(t) - \lambda(t) - \lambda_0(t) \log \left( \frac{\lambda_0(t)}{\lambda(t)} \right) \right] [Y_t - \mu_n(t)] dt \right|^{2k} \right] \\
&\leq \left( \int_0^T \left[ \lambda_0(t) - \lambda(t) - \lambda_0(t) \log \left( \frac{\lambda_0(t)}{\lambda(t)} \right) \right]^2 dt \right)^k \times \mathbb{E}_{\lambda_0}^{(n)} \left[ \left( \int_0^T [Y_t - \mu_n(t)]^2 dt \right)^k \right] \\
&\leq 2^{2k-1} C_{1k} n^k (A_{2k,1} + A_{2k,2}),
\end{aligned}$$

where, for  $\lambda \in B_{k,n}(\lambda_0; v_n, H)$ ,

$$\begin{aligned}
A_{2k,1} &:= \left[ \int_0^T \lambda_0^2(t) \log^2 \left( \frac{\lambda_0(t)}{\lambda(t)} \right) dt \right]^k \\
&\leq M_{\lambda_0}^{2k} \|\bar{\lambda}_0\|_\infty^k \left[ \int_0^T \bar{\lambda}_0(t) \log^2 \left( \frac{M_{\lambda_0} \bar{\lambda}_0(t)}{M_\lambda \bar{\lambda}(t)} \right) dt \right]^k \\
&\leq 2^{2k-1} M_{\lambda_0}^{2k} \|\bar{\lambda}_0\|_\infty^k \left[ E_2^k(\bar{\lambda}_0; \bar{\lambda}) + \left| \log \left( \frac{M_\lambda}{M_{\lambda_0}} \right) \right|^{2k} \right] \\
&\leq C \left[ E_2^k(\bar{\lambda}_0; \bar{\lambda}) + |M_\lambda - M_{\lambda_0}|^{2k} \right] \leq C v_n^{2k}
\end{aligned}$$

and

$$\begin{aligned}
A_{2k,2} &:= \left( \int_0^T [\lambda_0(t) - \lambda(t)]^2 dt \right)^k \\
&= \left( \int_0^T \{ (M_{\lambda_0} - M_\lambda) \bar{\lambda}_0(t) - M_\lambda [\bar{\lambda}(t) - \bar{\lambda}_0(t)] \}^2 dt \right)^k \\
&\leq 2^{2k-1} \|\bar{\lambda}_0\|_\infty^{2k} (M_{\lambda_0} - M_\lambda)^{2k} \\
&\quad + 2^{2k-1} M_\lambda^{2k} \left[ \int_0^T \left( \sqrt{\bar{\lambda}_0(t)} - \sqrt{\bar{\lambda}(t)} \right)^2 \left( \sqrt{\bar{\lambda}_0(t)} + \sqrt{\bar{\lambda}(t)} \right)^2 dt \right]^k \\
&\leq 2^{2k-1} \|\bar{\lambda}_0\|_\infty^{2k} (M_{\lambda_0} - M_\lambda)^{2k} + 2^k M_\lambda^{2k} (\|\bar{\lambda}_0\|_\infty + \|\bar{\lambda}\|_\infty)^k h^{2k}(\bar{\lambda}_0, \bar{\lambda}) \leq C v_n^{2k}.
\end{aligned}$$

Therefore,

$$A_{2k} \leq C(n v_n^2)^k.$$

To deal with  $B_{2k}$ , for any  $T > 0$ , we set

$$M_T := \int_0^T \log \left( \frac{\lambda_0(t)}{\lambda(t)} \right) [dN_t - Y_t \lambda_0(t) dt],$$

so  $(M_T)_T$  is a martingale. Using the Burkholder-Davis-Gundy Inequality (see Theorem B.15 in Karr (1991)), there exists a constant  $C(k)$  only depending on  $k$  such that, since

$2k > 1$ ,

$$\mathbb{E}_{\lambda_0}^{(n)} [|M_T|^{2k}] \leq C(k) \mathbb{E}_{\lambda_0}^{(n)} \left[ \left| \int_0^T \log^2 \left( \frac{\lambda_0(t)}{\lambda(t)} \right) dN_t \right|^k \right].$$

Therefore, for  $k > 1$ ,

$$\begin{aligned} B_{2k} &= \mathbb{E}_{\lambda_0}^{(n)} [|M_T|^{2k}] \\ &\leq 3^{k-1} C(k) \left( \mathbb{E}_{\lambda_0}^{(n)} \left[ \left| \int_0^T \log^2 \left( \frac{\lambda_0(t)}{\lambda(t)} \right) [dN_t - Y_t \lambda_0(t) dt] \right|^k \right. \right. \\ &\quad \left. \left. + \left| \int_0^T \log^2 \left( \frac{\lambda_0(t)}{\lambda(t)} \right) [Y_t - \mu_n(t)] \lambda_0(t) dt \right|^k \right. \right. \\ &\quad \left. \left. + \left| \int_0^T \log^2 \left( \frac{\lambda_0(t)}{\lambda(t)} \right) \mu_n(t) \lambda_0(t) dt \right|^k \right] \right) \\ &= 3^{k-1} C(k) (B_{k,2}^{(0)} + B_{k,2}^{(1)} + B_{k,2}^{(2)}), \end{aligned}$$

with

$$\begin{aligned} B_{k,2}^{(0)} &= \mathbb{E}_{\lambda_0}^{(n)} \left[ \left| \int_0^T \log^2 \left( \frac{\lambda_0(t)}{\lambda(t)} \right) [dN_t - Y_t \lambda_0(t) dt] \right|^k \right], \\ B_{k,2}^{(1)} &= \mathbb{E}_{\lambda_0}^{(n)} \left[ \left| \int_0^T \log^2 \left( \frac{\lambda_0(t)}{\lambda(t)} \right) [Y_t - \mu_n(t)] \lambda_0(t) dt \right|^k \right], \\ B_{k,2}^{(2)} &= \left| \int_0^T \log^2 \left( \frac{\lambda_0(t)}{\lambda(t)} \right) \mu_n(t) \lambda_0(t) dt \right|^k. \end{aligned}$$

This can be iterated: we set  $J = \min\{j \in \mathbb{N} : 2^j \geq k\}$  so that  $1 < k2^{1-J} \leq 2$ . There exists a constant  $C_k$ , only depending on  $k$ , such that for

$$B_{k2^{1-j}, 2^j}^{(1)} = \mathbb{E}_{\lambda_0}^{(n)} \left[ \left| \int_0^T \log^{2^j} \left( \frac{\lambda_0(t)}{\lambda(t)} \right) [Y_t - \mu_n(t)] \lambda_0(t) dt \right|^{k2^{1-j}} \right]$$

and

$$B_{k2^{1-j}, 2^j}^{(2)} = \left| \int_0^T \log^{2^j} \left( \frac{\lambda_0(t)}{\lambda(t)} \right) \mu_n(t) \lambda_0(t) dt \right|^{k2^{1-j}},$$



$$\begin{aligned}
B_{2k} &\leq C_k \left( \mathbb{E}_{\lambda_0}^{(n)} \left[ \left| \int_0^T \log^{2^J} \left( \frac{\lambda_0(t)}{\lambda(t)} \right) [dN_t - Y_t \lambda_0(t) dt] \right|^{k2^{1-J}} \right] \right. \\
&\quad \left. + \sum_{j=1}^J (B_{k2^{1-j}, 2^j}^{(1)} + B_{k2^{1-j}, 2^j}^{(2)}) \right) \\
&\leq C_k \left\{ \left( \mathbb{E}_{\lambda_0}^{(n)} \left[ \left| \int_0^T \log^{2^J} \left( \frac{\lambda_0(t)}{\lambda(t)} \right) [dN_t - Y_t \lambda_0(t) dt] \right|^2 \right] \right)^{k2^{-J}} \right. \\
&\quad \left. + \sum_{j=1}^J (B_{k2^{1-j}, 2^j}^{(1)} + B_{k2^{1-j}, 2^j}^{(2)}) \right\} \\
&= C_k \left[ \left( \mathbb{E}_{\lambda_0}^{(n)} \left[ \int_0^T \log^{2^{J+1}} \left( \frac{\lambda_0(t)}{\lambda(t)} \right) Y_t \lambda_0(t) dt \right] \right)^{k2^{-J}} + \sum_{j=1}^J (B_{k2^{1-j}, 2^j}^{(1)} + B_{k2^{1-j}, 2^j}^{(2)}) \right] \\
&= C_k \left[ \left( \int_0^T \log^{2^{J+1}} \left( \frac{\lambda_0(t)}{\lambda(t)} \right) \mu_n(t) \lambda_0(t) dt \right)^{k2^{-J}} + \sum_{j=1}^J (B_{k2^{1-j}, 2^j}^{(1)} + B_{k2^{1-j}, 2^j}^{(2)}) \right] \\
&= C_k \left[ B_{k2^{-J}, 2^{J+1}}^{(2)} + \sum_{j=1}^J (B_{k2^{1-j}, 2^j}^{(1)} + B_{k2^{1-j}, 2^j}^{(2)}) \right].
\end{aligned}$$

Note that, for any  $1 \leq j \leq J$ ,

$$\begin{aligned}
B_{k2^{1-j}, 2^j}^{(1)} &\leq \left[ \int_0^T \log^{2^{j+1}} \left( \frac{\lambda_0(t)}{\lambda(t)} \right) \lambda_0^2(t) dt \right]^{k2^{-j}} \times \mathbb{E}_{\lambda_0}^{(n)} \left[ \left( \int_0^T [Y_t - \mu_n(t)]^2 dt \right)^{k2^{-j}} \right] \\
&\leq C(M_{\lambda_0}^2 \|\bar{\lambda}_0\|_\infty)^{k2^{-j}} \left[ \int_0^T \log^{2^{j+1}} \left( \frac{M_{\lambda_0} \bar{\lambda}_0(t)}{M_\lambda \bar{\lambda}(t)} \right) \bar{\lambda}_0(t) dt \right]^{k2^{-j}} \times n^{k2^{-j}} \\
&\leq C \left[ \log^{2^{j+1}} \left( \frac{M_{\lambda_0}}{M_\lambda} \right) + E_{2^{j+1}}(\bar{\lambda}_0; \bar{\lambda}) \right]^{k2^{-j}} \times n^{k2^{-j}} \\
&\leq C(nv_n^2)^{k2^{-j}} \leq C(nv_n^2)^k,
\end{aligned}$$

where we have used (2.1). Similarly, for any  $j \geq 1$ ,

$$\begin{aligned}
B_{k2^{1-j}, 2^j}^{(2)} &\leq (nm_2 M_{\lambda_0})^{k2^{1-j}} \left[ \int_0^T \log^{2^j} \left( \frac{M_{\lambda_0} \bar{\lambda}_0(t)}{M_\lambda \bar{\lambda}(t)} \right) \bar{\lambda}_0(t) dt \right]^{k2^{1-j}} \\
&\leq C \left[ \log^{2^j} \left( \frac{M_{\lambda_0}}{M_\lambda} \right) + E_{2^j}(\bar{\lambda}_0; \bar{\lambda}) \right]^{k2^{1-j}} \times n^{k2^{1-j}} \leq C(nv_n^2)^{k2^{1-j}} \leq C(nv_n^2)^k.
\end{aligned}$$

Therefore, for any  $k > 1$ ,

$$V_{2k}(\lambda_0; \lambda) \leq \kappa(nv_n^2)^k,$$

where  $\kappa$  depends on  $C_{1k}$ ,  $k$ ,  $H$ ,  $\lambda_0$ ,  $m_1$  and  $m_2$ . Using previous computations, the case  $k = 1$  is straightforward. So, we obtain the result for  $V_k(\lambda_0; \lambda)$  for every  $k \geq 2$ .  $\square$

To prove Proposition 4.2, we use the following lemma whose proof is reported in Section 6.

**Lemma 5.1.** *Under condition (1.2), there exist constants  $\xi$ ,  $K > 0$ , only depending on  $M_{\lambda_0}$ ,  $\alpha$ ,  $m_1$  and  $m_2$ , such that, for any non-negative function  $\lambda_1$ , there exists a test  $\phi_{\lambda_1}$  so that*

$$\mathbb{E}_{\lambda_0}^{(n)}[\mathbf{1}_{\Gamma_n} \phi_{\lambda_1}] \leq 2 \exp(-Kn \|\lambda_1 - \lambda_0\|_1 \times \min\{\|\lambda_1 - \lambda_0\|_1, m_1\})$$

and

$$\sup_{\lambda: \|\lambda - \lambda_1\|_1 < \xi \|\lambda_1 - \lambda_0\|_1} \mathbb{E}_{\lambda}[\mathbf{1}_{\Gamma_n} (1 - \phi_{\lambda_1})] \leq 2 \exp(-Kn \|\lambda_1 - \lambda_0\|_1 \times \min\{\|\lambda_1 - \lambda_0\|_1, m_1\}).$$

*Proof of Proposition 4.2.* We consider the setting of Lemma 5.1 and a covering of  $S_{n,j}(v_n)$  with  $\mathbb{L}_1$ -balls of radius  $\xi j v_n$  and centers  $(\lambda_{l,j})_{l=1, \dots, D_j}$ , where  $D_j$  is the covering number of  $S_{n,j}(v_n)$  by such balls. We set  $\phi_{n,j} = \max_{l=1, \dots, D_j} \phi_{\lambda_{l,j}}$ , where the  $\phi_{\lambda_{l,j}}$ 's are defined in Lemma 5.1. So, there exists a constant  $\rho > 0$  such that

$$\mathbb{E}_{\lambda_0}^{(n)}[\mathbf{1}_{\Gamma_n} \phi_{n,j}] \leq 2D_j e^{-Knj^2 v_n^2} \text{ and } \sup_{\lambda \in S_{n,j}(v_n)} \mathbb{E}_{\lambda}^{(n)}[\mathbf{1}_{\Gamma_n} (1 - \phi_{n,j})] \leq 2e^{-Knj^2 v_n^2}, \quad \text{if } j \leq \frac{\rho}{v_n},$$

and

$$\mathbb{E}_{\lambda_0}^{(n)}[\mathbf{1}_{\Gamma_n} \phi_{n,j}] \leq 2D_j e^{-Knj v_n} \text{ and } \sup_{\lambda \in S_{n,j}(v_n)} \mathbb{E}_{\lambda}^{(n)}[\mathbf{1}_{\Gamma_n} (1 - \phi_{n,j})] \leq 2e^{-Knj v_n}, \quad \text{if } j > \frac{\rho}{v_n},$$

where  $K$  is a constant (see Lemma 5.1). We now bound  $D_j$ . First note that for any  $\lambda = M_{\lambda} \bar{\lambda}$  and  $\lambda' = M_{\lambda'} \bar{\lambda}'$ ,

$$\|\lambda - \lambda'\|_1 \leq M_{\lambda} \|\bar{\lambda} - \bar{\lambda}'\|_1 + |M_{\lambda} - M_{\lambda'}|. \quad (5.5)$$

Assume that  $M_{\lambda} \geq M_{\lambda_0}$ . Then,

$$\begin{aligned} \|\lambda - \lambda_0\|_1 &\geq \int_{\bar{\lambda} > \bar{\lambda}_0} [M_{\lambda} \bar{\lambda}(t) - M_{\lambda_0} \bar{\lambda}_0(t)] dt \\ &= M_{\lambda} \int_{\bar{\lambda} > \bar{\lambda}_0} [\bar{\lambda}(t) - \bar{\lambda}_0(t)] dt + (M_{\lambda} - M_{\lambda_0}) \int_{\bar{\lambda} > \bar{\lambda}_0} \bar{\lambda}_0(t) dt \\ &\geq M_{\lambda} \int_{\bar{\lambda} > \bar{\lambda}_0} [\bar{\lambda}(t) - \bar{\lambda}_0(t)] dt = \frac{M_{\lambda}}{2} \|\bar{\lambda} - \bar{\lambda}_0\|_1. \end{aligned}$$

Conversely, if  $M_{\lambda} < M_{\lambda_0}$ ,

$$\begin{aligned} \|\lambda - \lambda_0\|_1 &\geq \int_{\bar{\lambda}_0 > \bar{\lambda}} [M_{\lambda_0} \bar{\lambda}_0(t) - M_{\lambda} \bar{\lambda}(t)] dt \\ &\geq M_{\lambda_0} \int_{\bar{\lambda}_0 > \bar{\lambda}} [\bar{\lambda}_0(t) - \bar{\lambda}(t)] dt = \frac{M_{\lambda_0}}{2} \|\bar{\lambda} - \bar{\lambda}_0\|_1. \end{aligned}$$

So,  $2\|\lambda - \lambda_0\|_1 \geq (M_\lambda \vee M_{\lambda_0})\|\bar{\lambda} - \bar{\lambda}_0\|_1$  and we finally have

$$\|\lambda - \lambda_0\|_1 \geq \max \{ (M_\lambda \vee M_{\lambda_0})\|\bar{\lambda} - \bar{\lambda}_0\|_1/2, |M_\lambda - M_{\lambda_0}| \}. \quad (5.6)$$

So, for all  $\lambda = M_\lambda \bar{\lambda} \in S_{n,j}(v_n)$ ,

$$\|\bar{\lambda} - \bar{\lambda}_0\|_1 \leq \frac{2(j+1)v_n}{M_{\lambda_0}} \quad \text{and} \quad |M_\lambda - M_{\lambda_0}| \leq (j+1)v_n. \quad (5.7)$$

Therefore,  $S_{n,j}(v_n) \subseteq (\bar{S}_{n,j} \cap \mathcal{F}_n) \times \{M : |M - M_{\lambda_0}| \leq (j+1)v_n\}$  and any covering of  $(\bar{S}_{n,j} \cap \mathcal{F}_n) \times \{M : |M - M_{\lambda_0}| \leq (j+1)v_n\}$  will give a covering of  $S_{n,j}(v_n)$ . So, to bound  $D_j$ , we have to build a convenient covering of  $(\bar{S}_{n,j} \cap \mathcal{F}_n) \times \{M : |M - M_{\lambda_0}| \leq (j+1)v_n\}$ . We distinguish two cases.

- We assume that  $(j+1)v_n \leq 2M_{\lambda_0}$ . Then, (5.7) implies that  $M_\lambda \leq 3M_{\lambda_0}$ . Moreover, if

$$\|\bar{\lambda} - \bar{\lambda}'\|_1 \leq \frac{\xi j v_n}{3M_{\lambda_0} + 1} \quad \text{and} \quad |M_\lambda - M_{\lambda'}| \leq \frac{\xi j v_n}{3M_{\lambda_0} + 1},$$

then, by (5.5),

$$\|\lambda - \lambda'\|_1 \leq \frac{(M_\lambda + 1)\xi j v_n}{3M_{\lambda_0} + 1} \leq \xi j v_n.$$

By assumption (ii) of Theorem 2.1, this implies that, for any  $\delta > 0$ , there exists  $J_0$  such that for  $j \geq J_0$ ,

$$\begin{aligned} D_j &\leq D((3M_{\lambda_0} + 1)^{-1}\xi j v_n, \bar{S}_{n,j} \cap \mathcal{F}_n, \|\cdot\|_1) \times \left[ 2(j+1)v_n \times \frac{(3M_{\lambda_0} + 1)}{\xi j v_n} + \frac{1}{2} \right] \\ &\lesssim \exp(\delta(j+1)^2 n v_n^2). \end{aligned}$$

- We assume that  $(j+1)v_n > 2M_{\lambda_0}$ . If

$$\|\bar{\lambda} - \bar{\lambda}'\|_1 \leq \frac{\xi}{4} \quad \text{and} \quad |M_\lambda - M_{\lambda'}| \leq \frac{\xi(M_\lambda \vee M_{\lambda_0})}{4},$$

using again (5.5) and (5.7),

$$\|\lambda - \lambda'\|_1 \leq \frac{\xi M_\lambda}{4} + \frac{\xi(M_\lambda + M_{\lambda_0})}{4} \leq \frac{3\xi M_{\lambda_0}}{4} + \frac{\xi(j+1)v_n}{2} \leq \frac{7\xi(j+1)v_n}{8} \leq \xi j v_n,$$

for  $n$  large enough. By assumption (i) of Theorem 2.1, this implies that, for any  $\delta > 0$ ,

$$D_j \lesssim D(\xi/4, \mathcal{F}_n, \|\cdot\|_1) \times \log((j+1)v_n) \lesssim \log(jv_n) \exp(\delta n).$$

It is enough to choose  $\delta$  small enough to obtain the result of Proposition 4.2.  $\square$

## 6 Appendix

*Proof of Lemma 5.1.* For any  $\lambda$ , we denote by  $\mathbb{E}_{\lambda, \Gamma_n}^{(n)}[\cdot] = \mathbb{E}_\lambda^{(n)}[\mathbf{1}_{\Gamma_n} \times \cdot]$ . For any  $\lambda, \lambda'$ , we define

$$\|\lambda - \lambda'\|_{\tilde{\mu}_n} := \int_{\Omega} |\lambda(t) - \lambda'(t)| \tilde{\mu}_n(t) dt.$$

On  $\Gamma_n$  we have

$$m_1 \|\lambda - \lambda_0\|_1 \leq \|\lambda - \lambda_0\|_{\tilde{\mu}_n} \leq m_2 \|\lambda - \lambda_0\|_1. \quad (6.1)$$

The main tool for building convenient tests is Theorem 3 of Hansen et al. (2012) (and its proof) applied in the univariate setting. By mimicking the proof of this theorem from Inequality (7.5) to Inequality (7.7), if  $H$  is a deterministic function bounded by  $b$ , we have that, for any  $u \geq 0$ ,

$$\mathbb{P}_\lambda^{(n)} \left( \left| \int_0^T H_t (dN_t - d\Lambda_t) \right| \geq \sqrt{2vu} + \frac{bu}{3} \text{ and } \Gamma_n \right) \leq 2e^{-u}, \quad (6.2)$$

where we recall that  $\Lambda_t = \int_0^t Y_s \lambda(s) ds$  and  $v$  is a deterministic constant such that, on  $\Gamma_n$ ,  $\int_0^T H_t^2 Y_t \lambda(t) dt \leq v$  almost surely. For any non-negative function  $\lambda_1$ , we define the sets

$$A := \{t \in \Omega : \lambda_1(t) \geq \lambda_0(t)\} \quad \text{and} \quad A^c := \{t \in \Omega : \lambda_1(t) < \lambda_0(t)\}$$

and the following pseudo-metrics

$$d_A(\lambda_1, \lambda_0) := \int_A [\lambda_1(t) - \lambda_0(t)] \tilde{\mu}_n(t) dt \quad \text{and} \quad d_{A^c}(\lambda_1, \lambda_0) := \int_{A^c} [\lambda_0(t) - \lambda_1(t)] \tilde{\mu}_n(t) dt.$$

Note that  $\|\lambda_1 - \lambda_0\|_{\tilde{\mu}_n} = d_A(\lambda_1, \lambda_0) + d_{A^c}(\lambda_1, \lambda_0)$ . For  $u > 0$ , if  $d_A(\lambda_1, \lambda_0) \geq d_{A^c}(\lambda_1, \lambda_0)$ , define the test

$$\phi_{\lambda_1, A}(u) := \mathbf{1} \left\{ N(A) - \int_A \lambda_0(t) Y_t dt \geq \rho_n(u) \right\}, \quad \text{with } \rho_n(u) := \sqrt{2nv(\lambda_0)u} + \frac{u}{3},$$

where, for any non-negative function  $\lambda$ ,

$$v(\lambda) := (1 + \alpha) \int_{\Omega} \lambda(t) \tilde{\mu}_n(t) dt. \quad (6.3)$$

Similarly, if  $d_A(\lambda_1, \lambda_0) < d_{A^c}(\lambda_1, \lambda_0)$ , define

$$\phi_{\lambda_1, A^c}(u) := \mathbf{1} \left\{ N(A^c) - \int_{A^c} \lambda_0(t) Y_t dt \leq -\rho_n(u) \right\}.$$

Since for any non-negative function  $\lambda$ , on  $\Gamma_n$ , by (4.1),

$$(1 - \alpha) \int_{\Omega} \lambda(t) \tilde{\mu}_n(t) dt \leq \int_{\Omega} \lambda(t) \frac{Y_t}{n} dt \leq (1 + \alpha) \int_{\Omega} \lambda(t) \tilde{\mu}_n(t) dt, \quad (6.4)$$

inequality (6.2) applied with  $H = \mathbf{1}_A$  or  $H = \mathbf{1}_{A^c}$ ,  $b = 1$  and  $v = nv(\lambda_0)$  implies that, for any  $u > 0$ ,

$$\mathbb{E}_{\lambda_0, \Gamma_n}^{(n)}[\phi_{\lambda_1, A}(u)] \leq 2e^{-u} \quad \text{and} \quad \mathbb{E}_{\lambda_0, \Gamma_n}^{(n)}[\phi_{\lambda_1, A^c}(u)] \leq 2e^{-u}. \quad (6.5)$$

We now state a useful lemma whose proof is given below.

**Lemma 6.1.** *Assume condition (1.2) is verified. Let  $\lambda$  be a non-negative function. Assume that*

$$\|\lambda - \lambda_1\|_{\tilde{\mu}_n} \leq \frac{1 - \alpha}{4(1 + \alpha)} \|\lambda_1 - \lambda_0\|_{\tilde{\mu}_n}.$$

We set  $\tilde{M}_n(\lambda_0) = \int_{\Omega} \lambda_0(t) \tilde{\mu}_n(t) dt$  and we distinguish two cases.

1. Assume that  $d_A(\lambda_1, \lambda_0) \geq d_{A^c}(\lambda_1, \lambda_0)$ . Then,

$$\mathbb{E}_{\lambda, \Gamma_n}^{(n)}[1 - \phi_{\lambda_1, A}(u_A)] \leq 2 \exp(-u_A),$$

where

$$u_A = \begin{cases} u_{0A} n d_A^2(\lambda_1, \lambda_0), & \text{if } \|\lambda_1 - \lambda_0\|_{\tilde{\mu}_n} \leq 2\tilde{M}_n(\lambda_0), \\ u_{1A} n d_A(\lambda_1, \lambda_0), & \text{if } \|\lambda_1 - \lambda_0\|_{\tilde{\mu}_n} > 2\tilde{M}_n(\lambda_0), \end{cases}$$

and  $u_{0A}$ ,  $u_{1A}$  are two constants only depending on  $\alpha$ ,  $M_{\lambda_0}$ ,  $m_1$  and  $m_2$ .

2. Assume that  $d_A(\lambda_1, \lambda_0) < d_{A^c}(\lambda_1, \lambda_0)$ . Then,

$$\mathbb{E}_{\lambda, \Gamma_n}^{(n)}[1 - \phi_{\lambda_1, A^c}(u_{A^c})] \leq 2 \exp(-u_{A^c}),$$

where

$$u_{A^c} = \begin{cases} u_{0A^c} n d_{A^c}^2(\lambda_1, \lambda_0), & \text{if } \|\lambda_1 - \lambda_0\|_{\tilde{\mu}_n} \leq 2\tilde{M}_n(\lambda_0), \\ u_{1A^c} n d_{A^c}(\lambda_1, \lambda_0), & \text{if } \|\lambda_1 - \lambda_0\|_{\tilde{\mu}_n} > 2\tilde{M}_n(\lambda_0), \end{cases}$$

and  $u_{0A^c}$ ,  $u_{1A^c}$  are two constants only depending on  $\alpha$ ,  $M_{\lambda_0}$ ,  $m_1$  and  $m_2$ .

Note that, by (6.1), if  $d_A(\lambda_1, \lambda_0) \geq d_{A^c}(\lambda_1, \lambda_0)$ , by virtue of Lemma 6.1,

$$\begin{aligned} u_A &\geq \min\{u_{0A} n d_A^2(\lambda_1, \lambda_0), u_{1A} n d_A(\lambda_1, \lambda_0)\} \\ &\geq n d_A(\lambda_1, \lambda_0) \times \min\{u_{0A} d_A(\lambda_1, \lambda_0), u_{1A}\} \\ &\geq \frac{1}{2} n m_1 \|\lambda_1 - \lambda_0\|_1 \times \min\left\{\frac{1}{2} u_{0A} m_1 \|\lambda_1 - \lambda_0\|_1, u_{1A}\right\} \\ &\geq K_A n \|\lambda_1 - \lambda_0\|_1 \times \min\{\|\lambda_1 - \lambda_0\|_1, m_1\}, \end{aligned}$$

for  $K_A$  a positive constant small enough only depending on  $\alpha$ ,  $M_{\lambda_0}$ ,  $m_1$  and  $m_2$ . Similarly, if  $d_A(\lambda_1, \lambda_0) < d_{A^c}(\lambda_1, \lambda_0)$ ,

$$\begin{aligned} u_{A^c} &\geq \frac{1}{2} n m_1 \|\lambda_1 - \lambda_0\|_1 \times \min\left\{\frac{1}{2} u_{0A^c} m_1 \|\lambda_1 - \lambda_0\|_1, u_{1A^c}\right\} \\ &\geq K_{A^c} n \|\lambda_1 - \lambda_0\|_1 \times \min\{\|\lambda_1 - \lambda_0\|_1, m_1\}, \end{aligned}$$

for  $K_{A^c}$  a positive constant small enough only depending on  $\alpha$ ,  $M_{\lambda_0}$ ,  $m_1$  and  $m_2$ . Now, we set

$$\phi_{\lambda_1} = \phi_{\lambda_1, A}(u_A) \mathbf{1}_{\{d_A(\lambda_1, \lambda_0) \geq d_{A^c}(\lambda_1, \lambda_0)\}} + \phi_{\lambda_1, A^c}(u_{A^c}) \mathbf{1}_{\{d_A(\lambda_1, \lambda_0) < d_{A^c}(\lambda_1, \lambda_0)\}},$$

so that, with  $K = \min\{K_A, K_{A^c}\}$ , by using (6.5),

$$\begin{aligned} \mathbb{E}_{\lambda_0, \Gamma_n}^{(n)}[\phi_{\lambda_1}] &= \mathbb{E}_{\lambda_0, \Gamma_n}^{(n)}[\phi_{\lambda_1, A}(u_A) \mathbf{1}_{\{d_A(\lambda_1, \lambda_0) \geq d_{A^c}(\lambda_1, \lambda_0)\}}] \\ &\quad + \mathbb{E}_{\lambda_0, \Gamma_n}^{(n)}[\phi_{\lambda_1, A^c}(u_{A^c}) \mathbf{1}_{\{d_A(\lambda_1, \lambda_0) < d_{A^c}(\lambda_1, \lambda_0)\}}] \\ &\leq 2e^{-u_A} \mathbf{1}_{\{d_A(\lambda_1, \lambda_0) \geq d_{A^c}(\lambda_1, \lambda_0)\}} + 2e^{-u_{A^c}} \mathbf{1}_{\{d_A(\lambda_1, \lambda_0) < d_{A^c}(\lambda_1, \lambda_0)\}} \\ &\leq 2 \exp(-Kn \|\lambda_1 - \lambda_0\|_1 \times \min\{\|\lambda_1 - \lambda_0\|_1, m_1\}). \end{aligned}$$

If  $\|\lambda - \lambda_1\|_1 < \xi \|\lambda_1 - \lambda_0\|_1$ ,  $\xi = m_1(1 - \alpha)/[4m_2(1 + \alpha)]$ , then

$$\|\lambda - \lambda_1\|_{\tilde{\mu}_n} \leq \frac{1 - \alpha}{4(1 + \alpha)} \|\lambda_1 - \lambda_0\|_{\tilde{\mu}_n}$$

and Lemma 6.1 shows that

$$\begin{aligned} \mathbb{E}_{\lambda, \Gamma_n}^{(n)}[1 - \phi_{\lambda_1}] &\leq 2e^{-u_A} \mathbf{1}_{\{d_A(\lambda_1, \lambda_0) \geq d_{A^c}(\lambda_1, \lambda_0)\}} + 2e^{-u_{A^c}} \mathbf{1}_{\{d_A(\lambda_1, \lambda_0) < d_{A^c}(\lambda_1, \lambda_0)\}} \\ &\leq 2 \exp(-Kn \|\lambda_1 - \lambda_0\|_1 \times \min\{\|\lambda_1 - \lambda_0\|_1, m_1\}), \end{aligned}$$

which completes the proof of Lemma 5.1.  $\square$

*Proof of Lemma 6.1.* We only consider the case where  $d_A(\lambda_1, \lambda_0) \geq d_{A^c}(\lambda_1, \lambda_0)$ . The case  $d_A(\lambda_1, \lambda_0) < d_{A^c}(\lambda_1, \lambda_0)$  can be dealt with using similar arguments. So, we assume that  $d_A(\lambda_1, \lambda_0) \geq d_{A^c}(\lambda_1, \lambda_0)$ . On  $\Gamma_n$  we have

$$\begin{aligned} \int_A [\lambda_1(t) - \lambda_0(t)] Y_t dt &\geq n(1 - \alpha) \int_A [\lambda_1(t) - \lambda_0(t)] \tilde{\mu}_n(t) dt \\ &\geq \frac{n(1 - \alpha)}{2} \|\lambda_1 - \lambda_0\|_{\tilde{\mu}_n} \\ &\geq 2n(1 + \alpha) \|\lambda - \lambda_1\|_{\tilde{\mu}_n} \\ &\geq 2n(1 + \alpha) \int_A |\lambda(t) - \lambda_1(t)| \tilde{\mu}_n(t) dt \geq 2 \int_A |\lambda(t) - \lambda_1(t)| Y_t dt. \end{aligned}$$

Therefore,

$$\begin{aligned} \mathbb{E}_{\lambda, \Gamma_n}^{(n)}[1 - \phi_{\lambda_1, A}(u_A)] &= \mathbb{P}_{\lambda, \Gamma_n}^{(n)} \left( N(A) - \int_A \lambda(t) Y_t dt < \rho_n(u_A) + \int_A (\lambda_0 - \lambda)(t) Y_t dt \right) \\ &= \mathbb{P}_{\lambda, \Gamma_n}^{(n)} \left( N(A) - \int_A \lambda(t) Y_t dt < \rho_n(u_A) - \int_A (\lambda_1 - \lambda_0)(t) Y_t dt \right. \\ &\quad \left. + \int_A (\lambda_1 - \lambda)(t) Y_t dt \right) \\ &\leq \mathbb{P}_{\lambda, \Gamma_n}^{(n)} \left( N(A) - \int_A \lambda(t) Y_t dt < \rho_n(u_A) - \frac{1}{2} \int_A (\lambda_1 - \lambda_0)(t) Y_t dt \right). \end{aligned}$$

Assume that  $\|\lambda_1 - \lambda_0\|_{\tilde{\mu}_n} \leq 2\tilde{M}_n(\lambda_0)$ . This assumption implies that  $d_A(\lambda_1, \lambda_0) \leq \|\lambda_1 - \lambda_0\|_{\tilde{\mu}_n} \leq 2\tilde{M}_n(\lambda_0) \leq 2m_2M_{\lambda_0}$ . Since  $v(\lambda_0) = (1 + \alpha)\tilde{M}_n(\lambda_0)$ , with  $u_A = u_{0A}nd_A^2(\lambda_1, \lambda_0)$ , where  $u_{0A} \leq 1$  is a constant depending on  $\alpha$ ,  $m_1$  and  $m_2$  chosen later, we have

$$\rho_n(u_A) \leq nd_A(\lambda_1, \lambda_0)\sqrt{2u_{0A}(1 + \alpha)\tilde{M}_n(\lambda_0)} + \frac{u_{0A}nd_A^2(\lambda_1, \lambda_0)}{3} \leq K_1\sqrt{u_{0A}}nd_A(\lambda_1, \lambda_0)$$

as soon as  $K_1 \geq [2(1 + \alpha)\tilde{M}_n(\lambda_0)]^{1/2} + 2\tilde{M}_n(\lambda_0)\sqrt{u_{0A}}/3$ . Note that the definition of  $v(\lambda)$  in (6.3) gives

$$\begin{aligned} v(\lambda) &= (1 + \alpha) \int_{\Omega} \lambda_0(t)\tilde{\mu}_n(t)dt + (1 + \alpha) \int_{\Omega} [\lambda(t) - \lambda_0(t)]\tilde{\mu}_n(t)dt \\ &\leq v(\lambda_0) + (1 + \alpha)\|\lambda - \lambda_0\|_{\tilde{\mu}_n} \\ &\leq v(\lambda_0) + (1 + \alpha)[\|\lambda - \lambda_1\|_{\tilde{\mu}_n} + \|\lambda_1 - \lambda_0\|_{\tilde{\mu}_n}] \\ &\leq v(\lambda_0) + \frac{5 + 3\alpha}{4}\|\lambda_1 - \lambda_0\|_{\tilde{\mu}_n} \leq C_1, \end{aligned}$$

where  $C_1$  only depends on  $\alpha$ ,  $M_{\lambda_0}$ ,  $m_1$  and  $m_2$ . Combined with (6.4), this implies that, on  $\Gamma_n$ , if  $K_1 \leq (1 - \alpha)/[4\sqrt{u_{0A}}]$ , which is true for  $u_{0A}$  small enough,

$$\begin{aligned} \frac{1}{2} \int_A (\lambda_1 - \lambda_0)(t)Y_t dt - \rho_n(u_A) &\geq \frac{(1 - \alpha)n}{2}d_A(\lambda_1, \lambda_0) \left[ 1 - \frac{2K_1\sqrt{u_{0A}}}{1 - \alpha} \right] \\ &\geq \frac{(1 - \alpha)n}{4}d_A(\lambda_1, \lambda_0) \geq \sqrt{2nC_1r} + \frac{r}{3} \geq \sqrt{2nv(\lambda)r} + \frac{r}{3}, \end{aligned}$$

with

$$r = n \min \left\{ \frac{(1 - \alpha)^2}{128C_1}d_A^2(\lambda_1, \lambda_0), \frac{3(1 - \alpha)}{8}d_A(\lambda_1, \lambda_0) \right\}.$$

Inequality (6.2) then leads to

$$\mathbb{E}_{\lambda, \Gamma_n}^{(n)} [1 - \phi_{\lambda_1, A}(u_A)] \leq 2e^{-r}. \quad (6.6)$$

For  $u_{0A}$  small enough only depending on  $M_{\lambda_0}$ ,  $\alpha$ ,  $m_1$  and  $m_2$ , we have

$$\frac{(1 - \alpha)}{4\sqrt{u_{0A}}} \geq \sqrt{2(1 + \alpha)\tilde{M}_n(\lambda_0)} + \frac{2\tilde{M}_n(\lambda_0)\sqrt{u_{0A}}}{3}$$

so (6.6) is true. Since  $r \geq u_A$  for  $u_{0A}$  small enough, then

$$\mathbb{E}_{\lambda, \Gamma_n}^{(n)} [1 - \phi_{\lambda_1, A}(u)] \leq 2e^{-u_A}.$$

Assume that  $\|\lambda_1 - \lambda_0\|_{\tilde{\mu}_n} > 2\tilde{M}_n(\lambda_0)$ . We take  $u_A = u_{1A}nd_A(\lambda_1, \lambda_0)$ , where  $u_{1A} \leq 1$  is a constant depending on  $\alpha$  chosen later. We still consider the same test  $\phi_{\lambda_1, A}(u_A)$ .

Observe now that, since  $d_A(\lambda_1, \lambda_0) \geq \frac{1}{2}\|\lambda_1 - \lambda_0\|_{\tilde{\mu}_n} \geq \tilde{M}_n(\lambda_0)$ ,

$$\begin{aligned} \rho_n(u_A) &= \sqrt{2nu_A v(\lambda_0)} + \frac{u_A}{3} \\ &\leq n\sqrt{2(1+\alpha)u_{1A}\tilde{M}_n(\lambda_0)d_A(\lambda_1, \lambda_0)} + \frac{nu_{1A}}{3}d_A(\lambda_1, \lambda_0) \\ &\leq \left[ \sqrt{2(1+\alpha)} + \frac{1}{3} \right] n\sqrt{u_{1A}}d_A(\lambda_1, \lambda_0) \end{aligned}$$

and, under the assumptions of the lemma,

$$v(\lambda) \leq (1+\alpha)\tilde{M}_n(\lambda_0) + (1+\alpha)[\|\lambda - \lambda_1\|_{\tilde{\mu}_n} + \|\lambda_1 - \lambda_0\|_{\tilde{\mu}_n}] \leq C_2 d_A(\lambda_1, \lambda_0), \quad (6.7)$$

where  $C_2$  only depends on  $\alpha$ . Therefore,

$$\begin{aligned} \frac{1}{2} \int_A (\lambda_1 - \lambda_0)(t) Y_t dt - \rho_n(u_A) &\geq \frac{n(1-\alpha)}{2} \int_A [\lambda_1(t) - \lambda_0(t)] \tilde{\mu}_n(t) dt - \left( \sqrt{2(1+\alpha)} + \frac{1}{3} \right) \sqrt{u_{1A}} n d_A(\lambda_1, \lambda_0) \\ &\geq \left[ \frac{1-\alpha}{2} - \left( \sqrt{2(1+\alpha)} + \frac{1}{3} \right) \sqrt{u_{1A}} \right] n d_A(\lambda_1, \lambda_0) \\ &\geq \frac{1-\alpha}{4} n d_A(\lambda_1, \lambda_0), \end{aligned}$$

where the last inequality is true for  $u_{1A}$  small enough depending only on  $\alpha$ . Finally, using (6.7), since  $u_A = u_{1A} n d_A(\lambda_1, \lambda_0)$ , we have

$$\begin{aligned} \frac{1-\alpha}{4} n d_A(\lambda_1, \lambda_0) &\geq \sqrt{2nC_2 d_A(\lambda_1, \lambda_0) u_{1A} n d_A(\lambda_1, \lambda_0)} + \frac{1}{3} u_{1A} n d_A(\lambda_1, \lambda_0) \\ &\geq \sqrt{2nv(\lambda)u_A} + \frac{u_A}{3} \end{aligned}$$

for  $u_{1A}$  small enough depending only on  $\alpha$ . We then obtain

$$\mathbb{E}_{\lambda, \Gamma_n}^{(n)} [1 - \phi_{\lambda_1, A}(u_A)] \leq 2e^{-u_A},$$

which completes the proof.  $\square$

## References

- Aalen, O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.*, 6: 701–726.
- Andersen, P. K., Borgan, A., Gill, R. D., Keiding, N. (1993). Statistical models based on counting processes. Springer Series in Statistics. Springer-Verlag, New York.
- Belitser, E., Serra, P., van Zanten, J. H. (2013). Rate optimal Bayesian intensity smoothing for inhomogeneous Poisson processes. <http://arxiv.org/pdf/1304.6017v2>.



- Brunel, E., Comte, F. (2005). Penalized contrast estimation of density and hazard rate with censored data. *Sankhyā*, 3: 441–475.
- Brunel, E., Comte, F. (2008). Adaptive estimation of hazard rate with censored data. *Comm. Statist. Theory Methods*, 37: 1284–1305.
- Comte, F., Gaïffas, S., Guillaoux, A. (2011). Adaptive estimation of the conditional intensity of marker-dependent counting processes *Ann. Inst. Henri Poincaré Probab. Stat.*, 47: 1171–1196.
- Daley, D. J., Vere-Jones, D. (2003). An introduction to the theory of point processes. Vol. I, 2nd Edition. Probability and its Applications (New York). Springer-Verlag, New York, Elementary theory and methods.
- Daley, D. J., Vere-Jones, D. (2008). An introduction to the theory of point processes. Vol. II, 2nd Edition. Probability and its Applications (New York). Springer, New York, General theory and structure.
- de Boor, C. (1978). A practical guide to splines. Springer, New York.
- Gaïffas, S., Guillaoux, A. (2012). High-dimensional additive hazards models and the Lasso. *Electron. J. Stat.*, 6: 522–546.
- Ghosal, S., Ghosh, J. K., van der Vaart, A. W. (2000). Convergence rates of posterior distributions. *Ann. Statist.*, 28: 500–531.
- Ghosal, S., van der Vaart, A. (2007). Convergence rates of posterior distributions for noniid observations. *Ann. Statist.*, 35: 192–223.
- Groeneboom, P. (1985). Estimating a monotone density. In: Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer. Wadsworth Statist./ Prob. Ser., Belmont, CA Wadsworth, pp. 539–555.
- Hansen, N. R., Reynaud-Bouret, P., Rivoirard, V. (2012). Lasso and probabilistic inequalities for multivariate point processes. *arXiv:1208.0570*.
- Härdle, W., Kerkycharian, G., Picard, D., Tsybakov, A. (1998). Wavelets, approximation, and statistical applications. Vol. 129 of Lecture Notes in Statistics. Springer-Verlag, New York.
- Ishwaran, H., James, L. F. (2004). Computational methods for multiplicative intensity models using weighted gamma processes: proportional hazards, marked point processes, and panel count data. *Journal of the American Statistical Association*, 99, 175–190.
- Karr, A. F. (1991). Point processes and their statistical inference, 2nd Edition. Vol. 7 of Probability: Pure and Applied. Marcel Dekker Inc., New York.
- Kottas, A., Sansó, B. (2007). Bayesian mixture modeling for spatial Poisson process intensities, with applications to extreme value analysis. *J. Statist. Plann. Inference*, 137, 3151–3163.

- Kuo, L., Ghosh, S. K. (1997). Bayesian nonparametric inference for nonhomogeneous Poisson processes. Tech. rep., University of Connecticut, Department of Statistics.
- Lo, A., Weng, C.-S. (1989). On a class of Bayesian nonparametric estimates: II. Hazard rate estimates. *Annals of the Institute of Statistical Mathematics*, 41: 227–245.
- Lo, A. Y. (1982). Bayesian nonparametric statistical inference for Poisson point processes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 59: 55–66.
- Reynaud-Bouret, P. (2006). Penalized projection estimators of the Aalen multiplicative intensity. *Bernoulli*, 12: 633–661.
- Rivoirard, V., Rousseau, J. (2012). Posterior concentration rates for infinite dimensional exponential families. *Bayesian Analysis*, 7: 311–334.
- Salomond, J. B. (2013). Concentration rate and consistency of the posterior under monotonicity constraints. Tech. rep., Univ. Paris Dauphine.
- Williamson, R. E. (1956). Multiply monotone functions and their Laplace transforms. *Duke Math. J.*, 23: 189–207.
- Wong, W. H., Shen, X. (1995). Probability inequalities for likelihood ratios and convergence rates of sieve MLEs. *Ann. Statist.*, 23: 339–362.